us-08-482-402a-2.rge

[handwritten: 3/3/04]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

On nucleic - nucleic search, using sw model

Run on: November 26, 2001, 09:16:10 ; Search time 2192.46 Seconds
           (without alignments)
           23115.284 Million cell updates/sec

Title: US-08-482-402A-2
Perfect score: 3072
Sequence: 1 gagycaattgagycgccat..........atagctgcaaaaaaaaaaa 3072

Scoring table: IDENTITY_NUC
                Gapop 10.0 ; Gapext 1.0

Searched: 1472140 seqs. 8248589755 residues

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:
        1: gb_ba.*
        2: gb_htg.*
        3: gb_inv.*
        4: gb_om.*
        5: gb_ov.*
        6: gb_pat.*
        7: gb_ph.*
        8: gb_pl.*
        9: gb_pr.*
       10: gb_ro.*
       11: gb_sts.*
       12: gb_ba2.*
       13: gb_un.*
       14: gb_vi.*
       15: gb_vrt.*
       16: em_fun.*
       17: em_hum.*
       18: em_inv.*
       19: em_om.*
       20: em_or.*
       21: em_ph.*
       22: em_pl.*
       23: em_pr.*
       24: em_ro.*
       25: em_sts.*
       26: em_sts.*
       27: em_vrl.*
       28: em_vi.*
       29: em_vi.*
       30: em_htgo_hum.*
       31: em_htgo_rod.*
       32: em_htgo_rod.*
       33: em_htg_inv.*
       34: em_htg_hum.*
       35: em_htg_rod.*
       36: em_htg_other.*

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3060	99.6	3060	9	HUMTPOC	M17755 Homo sapien
2	3004	97.8	3048	9		AB029419 Sequence
3	3004	97.8	3048	9	HUMTPOA	J02969 Human thyro
4	2991.4	97.4	3027	9	HSTPO	Y00406 Human mRNA
5	2877	93.2	2877	9		A92132 Sequence 1
6	2652	86.3	2877	9	HUMTPOB	J02970 Human thyro
7	2649.6	86.2	2845	9	HSASTPO2	X17358 Human mRNA
8	2645	86.5	2845	9		S22650 Antigen fo
9	1655.2	53.9	3237	10	RXYPO	X17796 Rat mRNA fo
10	1628.6	53.0	3291	10	MMTPER	X60703 Mus muscula
11	1613.4	52.4	3291	10	RATTPOPER	M63556 Rat thyroid
12	1370	31.6	1046	9	RXYPO	X17796 Porcine mRN
13	516	16.8	1996	9	HUMTPO07	M25707 Human thyro
14	510	16.6	16016	10	HUMO408011	AC060811 Homo sapi
15	510.4	15.4	1142	4	SSTPOR1	X04645 Porcine mRN
16	469.8	15.3	2238	9	S56200	S56200 myeloperox
17	469.8	15.3	2568	9	HUMOR2	A08002 H.sapiens m
18	469.8	15.3	2568	9	HUMOR2	J02694 Human myelo
19	469.8	15.3	3213	9	HSMPOR	X04876 Human mRNA
20	469.8	15.3	3215	9	E12629	E12629 cDNA encodi
21	461.4	15.0	3215	9	MISEPO	M19507 Human mRNA
22	461.4	15.0	2666	10	MUSEPO	L77979 Mus muscula
23	461.4	15.0	2694	10	MUSEPO	D13553 Mouse mRNA
24	455.2	14.8	2945	10	MMMPO	D13550 Mouse mRNA
25	438.2	14.3	2558	9	HSEP	X14346 human mRNA
26	433.6	14.1	2915	9	AB022196	AB022196 Clona int
27	417	13.6	3861	9	AF343034	AF343034 Ovis rer
28	417	13.6	5510	9	D86983	D86983 Human mRNA
29	417	13.6	6847	9	AF200348	AF200348 Homo sapi
30	411	13.4	3788	9	AF378814	AF378814 Danio rer
31	397.4	12.9	2768	9	HSU39573	U39573 Human saliv
32	397	12.9	3179	9	AB022197	AB022197 Halocynth
33	386	12.6	2841	9	AB028841	AB028841 Branchiost
34	386	12.6	2710	9	BOVLPO	M58150 Bovine lact
35	379.6	12.4	2740	4	AF027970	AF027970 Ovis arie
36	379.6	12.4	2734	9	OAU19675	U19675 Camelus dr
c 37	348.8	11.7	17245	9	AC009471	AC009471 Homo sapi
38	338.4	11.0	435	9	S64243	S64243 thyroid per
c 39	310	10.1	30025	9	AC026475	AC026475 Drosophil
c 40	305	9.9	80254	9	AC015531	AC015531 Drosophil
c 41	301.8	9.8	2512	9	U68724	U68724 Xenopus lae
42	285	9.3	3611	9	HDU11052	U11052 Human lacto
43	289.4	9.4	1399	9	HUMLPO	M58151 Human lacto
44	265	8.6	1328	9	HUMTPO08	M25708 Human thyro
45	261.2	8.5	1448	9	AF364820	AF364820 Xenopus 1

ALIGNMENTS

RESULT   1
HUMTPOC
LOCUS   HUMTPOC   3060 bp   mRNA   PRI   26-APR-1999
DEFINITION Homo sapiens thyroid peroxidase (TPO) mRNA, complete cds.
ACCESSION   M17755
VERSION   M17755.2 GI:4680720
KEYWORDS   .
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 732 to 1573)
AUTHORS   Seto,P., Magnusson,R.P., Gestautas,J., Portmann,L.,
            DeGroot,L.J. and Rapoport,B.
TITLE   Isolation of a complementary DNA clone for thyroid microsomal
            antigen. Homology with the gene for thyroid peroxidase
JOURNAL   J. Clin. Invest. 80 (4), 1205-1208 (1987)
MEDLINE   88008367
REFERENCE   2 (bases 1 to 3060)
AUTHORS   Magnusson,R.P., Chazenbalk,G.D., Gestautas,J., Seto,P., Filetti,S.,

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

FEATURES   Location/Qualifiers

source   1..3060
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="2"
            /clone="PM-5"
            /tissue_type="thyroid"
            /note="isolated from patient with Graves' disease"

gene   1..3060
            /gene="TPO"

CDS   ...
            /gene="TPO"
            /function="primary enzyme for thyroid hormone synthesis"
            /note="similar to Homo sapiens thyroid peroxidase"
            /product="thyroperoxidase; membrane-associated cell surface protein;
            primary autoantigen in human autoimmune thyroiditis;
            microsomal thyroid autoantigen; thyroid microsomal antigen"
            /codon_start=1
            /product="thyroid peroxidase"
            /db_xref="GI:4680721"

misc_feature 2626..2697
            /gene="TPO"
            /note="putative transmembrane-region site"

BASE COUNT
ORIGIN

```
Qy     1 gagcaatgagaacgggcccttcgaggacgttctgaaaatactcagcagtgca      60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 gagcaatgagaacgggcccttcgaggacgttctgaaaatactcagcagtgca      60

Qy    61 gttggctgcagaggaggaaaaaaagatgagagcggcgtgctgtgctgcagcggcgtt    120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
db    61 gttggctgcagaggaggaaaaaaagatgagagcggcgtgctgtgctgcagcggcgtt    120
```

(Sequence alignment continues for db/Qy pairs through position 1200)

```
Db   1141  CACGAGCCGCCTCCGGGACTCCGGGCGCCGCGCCTTACCTGCCCTTTCGTGCCGCCACGCGGCGCCT  1200
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1201  gcggcctgctgccgccgcgccaggtccccggggagaagaccgcggccggctcctggccg... 1260
Db   1201  GCGGCCTGTGCGCCAGCCCGGCAACCCGAGAGAACCCCGGACCCCCGCTTCCTGCCC  1260
...
```

(The body of this page consists of a DNA sequence alignment between query "Qy" and database "Db" sequences, running from position 1141 to 3060, followed by the result record below.)

```
Query Match      97.8%;   Score 3004;  DB 6;  Length 3048;
Best Local Similarity   99.3%;   Pred. No. 0;
Matches 3028;  Conservative   0;  Mismatches  15;  Indels   5;  Gaps  1;
```

```
Qy    18  cattcaagagaattacagcctgaaaattacctcagcagtgagttggctgagaaagaagga     77      Qy  1033  atgaacggttgaactcttcctgacgcgtccaccgtgtatggccgctccagcccccgcccta  1092
db     1  CATTTCAGAGAGTTACAGCCTGAAAATTACTCAGCAGTGAGTTGGCTGAGAAAGAAGGA     60      db  1021  ATGAACGGTTGAACTCTTCCTTCCTGGACGCGTCCACCGTGTATGGCCAGCTCCCCGGCCCTA  1080

Qy    78  aaaa-----agaatgagacgtcgctctgctctataccgtcagtatggcctgcaca      132      Qy  1093  gagaggcagctcgcggaactggacagctcctcccctcccccgcgttccacgcgacgtctc  1152
db     1  AAAAAGGTCAGAATGAGACGTCGCTCTGCTCTATACCGTCACGCTGGTTATGGCCTGCACA  120      db  1081  GAGAGGCAGCTCGCGGAACTGGACAGCTGCCTCAGCAGGCTCCCGCGTTCCACGCGACGTCTC  1140

Qy   133  gaagccttcctccccttcatcctggaagggaaagaacttccttgggcaggagaggcgag  192      Qy  1153  cggggatccggcgaaacctggagaaccccgtgctcgtcggccgctccgccgtctccggcc  1212
db   121  GAAGCCTTCCTCCCCTTCATCCTGGAGGAGAGAAGGAAGTACCTCTTTGGGCAGGAGAGGAG  180      db  1141  CGGGATCCGGCGCGGAACCTGGAGAACCCCGTGCTCGTCGGCCGCCTCTCCGGGTGCGG  1200

Qy   193  tctcagtccctctctctgggagacagagggctggtgacccgcctgatacgcc      252      Qy  1213  ccgagccgggagaaacctccggaggacccgtgctcctggccgagacaggagaggaagcgc  1272
db   181  TCTCGTGTCGTCTAGGGTTGTGAGGAAAGACAGACTCCCTGCTGGAGCACCGCCTGATACAGCC  240      db  1201  CCCGAGCCGGGCAGCAATCCCCGGAGGACCCGTGCTTCTTGGCCGGAGACAGGACGAGCGCC  1260

Qy   253  acgatgcagagaaacctcaagaaagaggtcctcttcgagctagctctgctctttt      312      Qy  1273  gccagcgagctgcacgcacgcctgctcagccggcgcaaccgc  1332
db   241  ACGATGCAGAGAAACCTCAAGAAAGAGGATTCCTTTCGCGACTTAGCTGCTGTCTTTT  300      db  1261  GCCAGCGAGCTGCACGCACGCCCTGGTCCTGTCGCGCCGACAACCGC  1320

Qy   313  tccaaactccctgacccagggagagtgatgataagataagaaca      372      Qy  1333  ctgcctcggctgtccaacgtgtctccaacagcgccctcggcccctcggcctacgcaacg  1392
db   301  TCCAAACTTCCTGACCCAACAGAGGTGATTGATAAGAGATAATTAGAAACA  360      db  1321  CTGCCTCGGCTGTCCAACGTGTCTCCAACAGCGCCCTCGGCGCCCTCGGCCTACGACCAGG  1380

Qy   373  tcaatacagcctgaaaagaaaactcaacctgaaactcacacacatacagcc      432      Qy  1393  gggggaagacctgtggggctctgaccagggaaggatacaaccagg  1452
db   361  TCAATAAACAGCGATGAAAAGAAAAGTCAACTCAACAATTACACACACATACAACCAAG  420      db  1381  GCGGGAAGCGCCTGTGGCGGGCTCTGACCAGGAATCACCCTGGAGGGATTACATCCCCAGG  1440

Qy   433  gatgctttatcagagagctctgatgatcattcccaaatactctcggatgtctccttac  492      Qy  1453  ccactacaggtgctccaacgtgttctccaacagcagccgctccggctcaccagcgaccgc  1512
db   421  GATGCTTTATCAGAGAGCTCTGCTGCGACATCACAATCACAGCGTCGAGATGTCTCCCTTAC  480      db  1441  ATCCTGGAGACCCAGGCTTCCAACGTCCCTGGTGTGCTCGCTGATACGAGCTCCACTCTCACC  1500

Qy   493  atgctgcccccaaatgcccaaacactttgcctgcagcaactacagccatacagctacacaga  552      Qy  1513  gccaaccagtgtctccaacgtgctctcaaacagcgccctctggccctcggccacgacagc  1572
db   481  ATGCTGCCCCCAAATGCCCAAACACTTTGCCTGCTGCAACTACGCATACTACAGCTACACAGA  540      db  1501  GCCAAACCAGTGTCTCAAACGTGCTCTCCAACAGCGCCCTCCGGCCCTCGGCCACGACAGC  1560

Qy   553  gctgcaacaacgcccacacagtggctcacacagtggctctctcgacctacctgacacactg  612      Qy  1573  ctcccccggctgtccaggagggtgcaacagcacctcctggcaggctcggggtcgttggac  1632
db   541  GCTTGCAACAACAGCACCACCCCAGAGTGGCTCACACACTTAGTCGGCTCGACTGTCCCAGG  600      db  1561  CTCCCCTCGGCTGTGCAGGAGCTCTGACAACACCCTCGATCCGCCCTCGGGTCGTTCAGG  1620

Qy   613  ccccctccaatctatagaagggctcagctcaagggagaaagccccggatctctg  672      Qy  1633  cgtggctcgacaacaagcacgctcatcatcaccggaagattacgggtggtggtttggac  1692
db   601  CTCCCCCAGTCTATGAGAGGGCCTCTGCTCTCAGACCTGGAACCCCCGGAGCTCTCTG  660      db  1621  CGTGGCTGGACAACAAGCACGGTCGATCATCAACCGCCTGCGAGTGGTTTGGAC  1680

Qy   673  tacaacgggttccacctccccccgggatgatcattcaagttca  732      Qy  1693  caactacaggtgctccttcttctgcaagaccgccaaactcaggctcggggatcgcagtctg  1752
db   661  TACAACGGGTTCCACCTGCCCGCCCGGGCTCGCTGGACGTGATGTCCATTTAC  720      db  1681  CCACTACAGGTGCTCCTTCTTGCGCAAGACCGCCAACCTGGAGATCGCCAGCTGATG  1740

Qy   733  aatgagttgtcacagagatgacccgatcctaatctgacctgatggggacaatac  792      Qy  1753  aacgagagcctggaaagccttcttgtgctgaccatcaggcacctcggagttca  1812
db   721  AATGAGTTGTCACACAGATGACCCTGATCCTAATCTGACCTCGATGGGGACAATAC  780      db  1741  AACGAGAGCCTGGACAGGTGCCGAACCTCGTGCCAGTCCAATTTCGAGATCGCGG  1800

Qy   793  atcgaccaagcatatctctccaacaaagacaagaacaccgtgttttctctggggtgagg  852      Qy  1813  tccatccacgcccgactcgaatggcggcaccgggagaacatgcaac  1872
db   781  ATCGACCACAGATGACCTTCACAGAACAAGAACACCGTGTTTTCTGGGGGAGG  840      db  1801  TCCATCAACGTCCAAGCTGGACAGGCTGCAGGTTCACAAGTACAATGAGTGCAGGAG  1860

Qy   853  tctgaccgcaagaaattctggacttcatagacgtcctgccaacaagctgctctgagctgg  912      Qy  1873  tctgcggctcacagaaggaacatcatcccgaacatcac  1932
db   841  GCTGACTGCCAGATGAACTTCTGCAGAGTACAACAGCACCAAGCTGCTCTGAGCTGG  900      db  1861  TCCATCAACCTCACGGCCCGCAGCACCGCAGGTCACTCCAGCAAGCAGGGCTGG  1920

Qy   913  gaagcccgcacagccggccaggcacgcctgctcaccgcccgcgcgc  972      Qy  1933  acgttggccaaagatctggacttgtagaagatcctcgacagcacctgatcgtctgg  1992
db   901  GAAGGCCCCAGCCGCAGCCGGCCAGGCACGGCCTGCTCACCCGCCCGCCGCAGCAG  960      db  1921  ACGGTGGCCAACGAGATCTTGGACTTGTGTCGCAGGACGATCCAGCTGATCGTCTCGG  1980

Qy   973  gcgcccggagcagagggagctccttcttcttgggaacctctccagctccggagca  1032    Qy  1993  gagcctgagcgaaaactccctcccagggctgaagacctgtgcttgctgtctc  2052
db   961  GCCACAGGAACCAAGGGCGGGCCTCTTCTTGGCCGAACCTCTCCAGCTCCGGAGCAG  1020    db  1981  GGAGGCTTAGCTGAAAACTCCCGAAGGCTCGGACAGGCCCTGTTGCTGTCTC  2040

                                                                    Qy  2053  attgaagaacatgaggcgcgcaggcagcgagcagg  2112
                                                                    db  2041  ATTGAGAAGACAGTGAGGCGCGTGGTTTGGGCGGACGAAACAGCCAC  2100
```

ORGANISM   mRNA, clone pHPPO-2.8.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 3048)
AUTHORS   Kimura,S., Kotani,T., McBride,O.W., Umeki,K., Hirai,K., Nakayama,T.
            and Ohtaki,S.
TITLE   Human thyroid peroxidase: complete cDNA and protein sequence,
            chromosome mapping, and identification of two alternately spliced
            mRNAs
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 84 (16), 5555-5559 (1987)
MEDLINE   87289647
COMMENT   Draft entry and computer-readable sequence for [1] kindly provided
            by S.Kimura 09-SEP-1991.

FEATURES   Location/Qualifiers
     source   1..3048
                     /organism="Homo sapiens"
                     /chromosome="2"
                     /map="2pter-p24"
     mRNA   <1..3048
                     /gene="TPO"
     gene   73..2874
                     /gene="TPO"
     CDS   /gene="TPO"
                     /note="thyroid peroxidase"
                     /codon_start=1
                     /db_xref="GDB:G00-120-446"
                     /protein_id="AAA61215.1"
                     /db_xref="GI:339867"
                     /translation="MRALAVLSVTLVMACTEAFFPISRGKILKWLKKPEESRVSVLE
                     ESKRLVDTAYMPYLKDSLMSNTQQDSLASGRIASGGREKSLQALSLPCRTSMKITRQ
                     ..."

BASE COUNT   671 a   ...   566 t
ORIGIN
            Chromosome 2pter-q11.

Query Match   97.8%;   Pred. No. 0;
Best Local Similarity   99.3%; Score 3004; Length 3048;
Matches 3028; Conservative 0; Mismatches 15; Indels 5; Gaps 1;

Qy ... gtctcagtcacgcagtacgaccg ...
        |||||||||||||||||||||||
Db ... gtctcagtcacgcagtacgaccg ...

RESULT   3
LOCUS   HUMTPOA   3048 bp   mRNA   PRI   14-JAN-1995
DEFINITION   Human thyroid peroxidase mRNA, clone pHPPO-2.8.
ACCESSION   J02969
VERSION   J02969.1 GI:339866
KEYWORDS   alternative splicing; thyroid peroxidase.
SOURCE   Human thyroid gland (from patients with Grave's disease), cDNA to

```
Db    241 ACGATGCAGAGAAACCTCAAGAAAAGAGGAATCCTTTCTCCAGCTCAGCTTCTGTCTTTT 300

Qy    313 tccaaacttcctgagccaacaagcggagtgattgcccgagcagcagagataatggaaaca 372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCCAAACTTCCTGAGCCAACAAGCGGAGTGATTGCCCGAGCAGCAGAGATAATGGAAACA 360

Qy    373 tcaatacaagcgatgaaaagaaaagtcaacctgaaaactcaacaatcacagcatccaacg 432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCAATACAAGCGATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACG 420

Qy    433 gatgctttatcagaagatctgctgagcatcattgcaaacatgtctggatgtctcccttac 492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GATGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTCTCCCTTAC 480

Qy    493 atgctgcccccaaaatgcccaaacacttgcctggcgaacaaatacaggcccatcacagga 552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATGCTGCCCCCAAAATGCCCAAACACTTGCCTGGCGAACAAATACAGGCCCATCACAGGA 540

Qy    553 gcttgcaacaacagagaccaccccagatggggcgcctccaacacggccctggcacgatgg 612
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GCTTGCAACAACAGAGACCACCCCAGATGGGGCGCCTCCAACACGGCCCTGGCACGATGG 600

Qy    613 ctccctccagtctatgaggacggcttcagtcagccccgaggctggaaccccggcttcttg 672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTCCCTCCAGTCTATGAGGACGGCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTG 660

Qy    673 tacaacgggttcccactgcccccggtccgggaggtgacaagacatgtcattcaagtttca 732
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TACAACGGGTTCCCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCA 720

Qy    733 aatgaggttgtcacagatgatgaccgctattctgacctcctgatggcatggggacaatac 792
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AATGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGGGACAATAC 780

Qy    793 atcgaccacgacatcgcgttcacaccacagagcaccagcaaagctgccttcgggggaggg 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAAGCTGCCTTCGGGGGAGGG 840

Qy    853 tctgactgccagatgacttgtgagaaccaaaacccatgttttcccatacaactcccggag 912
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GCTGACTGCCAGATGACTTGTGAGAACCAAAACCCATGTTTTCCCATACAACTCCCGGAG 900

Qy    913 gaggcccggccggccgcgggcaccgcctgtctgcccttctaccgctcttcggccgcctgc 972
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GAGGCCCGGCCGGCCGCGGGCACCGCCTGTCTGCCCTTCTACCGCTCTTCGGCCGCCTGC 960

Qy    973 ggcaccggggaccaaggcgcgctctttgggaacctgtccacggccaacccgaggcagcag 1032
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    961 GGCACCGGGGACCAAGGCGCGCTCTTTGGGAACCTGTCCACGGCCAACCCGCGGCAGCAG 1020

Qy   1033 atgaacgggttgacctcgttcctggacgcgtccaccgtgtatggcagctccccggcccta 1092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ATGAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTGTATGGCAGCTCCCCGGCCCTA 1080

Qy   1093 gagaggcagctgcggaactggaccagtgccgaagggctgctccgcgtccacggccgcctc 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db   1081 GAGAGGCAGCTGCGGAACTGGACCAGTGCCGAAGGGCTGCTCCGCGTCCACGCGCGCCTC 1140

Qy   1153 cgggactccggccgcgcctacctgcccttcgtgccgccacgcgcgcctgcggcctgtgcg 1212
          |||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db   1141 CGGGACTCCGGCCGCGCCTACCTGCCCTTCGTGCCGCCACGGCGGCCTGCGGCCTGTGCG 1200

Qy   1213 cccgagcccggcaaccccggagagacccgcgggccctgcttcctggccggagacggccgc 1272
          ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CCCGAGCCCGGCATCCCCGGAGAGACCCGCGGGCCCTGCTTCCTGGCCGGAGACGGCCGC 1260

Qy   1273 gccagcgaggtcccctccctgacggcactgcacacgctgtggctgcgcgagcacaaccgc 1332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GCCAGCGAGGTCCCCTCCCTGACGGCACTGCACACGCTGTGGCTGCGCGAGCACAACCGC 1320

Qy   1333 ctggccgcggcgctcaaggccctcaatgcgcactggagcgcggacgccgtgtaccaggag 1392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTGGCCGCGGCGCTCAAGGCCCTCAATGCGCACTGGAGCGCGGACGCCGTGTACCAGGAG 1380
```

```
Qy   1393 gcgcgcaaggtcgtgggcgctctgcaccagatcatcaccctgagggattacatccccagg 1452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GCGCGCAAGGTCGTGGGCGCTCTGCACCAGATCATCACCCTGAGGGATTACATCCCCAGG 1440

Qy   1453 atcctgggacccgaggccttccagcagtacgtgggtccctatgaaggctatgactccacc 1512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ATCCTGGGACCCGAGGCCTTCCAGCAGTACGTGGGTCCCTATGAAGGCTATGACTCCACC 1500

Qy   1513 gccaaccccactgtgtccaacgtgttctccacagccgccttccgcttcggccatgccacg 1572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCCAACCCCACTGTGTCCAACGTGTTCTCCACAGCCGCCTTCCGCTTCGGCCATGCCACG 1560

Qy   1573 atccacccgctggtgaggaggctggacgccagcttccaggagcaccccgacctgcccggg 1632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ATCCACCCGCTGGTGAGGAGGCTGGACGCCAGCTTCCAGGAGCACCCCGACCTGCCCGGG 1620

Qy   1633 ctgtggctgcaccaggctttcttcagcccatggacattactccgtggaggtggtttggac 1692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CTGTGGCTGCACCAGGCTTTCTTCAGCCCATGGACATTACTCCGTGGAGGTGGTTTGGAC 1680

Qy   1693 ccactaatacgaggccttcttgcaagaccagccaaactgcaggtgcaggatcagctgatg 1752
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CCACTAATACGAGGCCTTCTTGCAAGACCAGCCAAACTGCAGGTGCAGGATCAGCTGATG 1740

Qy   1753 aacgaggagctgacggaaaggctctttgtgctgtccaattccagcaccttggatctggcg 1812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AACGAGGAGCTGACGGAAAGGCTCTTTGTGCTGTCCAATTCCAGCACCTTGGATCTGGCG 1800

Qy   1813 tccatcaacctgcagaggggccgggaccacgggctgccaggttacaatgagtggagggag 1872
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TCCATCAACCTGCAGAGGGGCCGGGACCACGGGCTGCCAGGTTACAATGAGTGGAGGGAG 1860

Qy   1873 ttctgcggcctgcctcgcctggagacccccgctgacctgagcacagccatcgccagcagg 1932
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TTCTGCGGCCTGCCTCGCCTGGAGACCCCCGCTGACCTGAGCACAGCCATCGCCAGCAGG 1920

Qy   1933 agcgtggccgacaagatcctggacttgtacaagcatcctgacaacatcgatgtctggctg 1992
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 AGCGTGGCCGACAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATGTCTGGCTG 1980

Qy   1993 ggaggcttagctgaaaacttcctccccagggctcggacagggcccctgtttgcctgtctc 2052
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GGAGGCTTAGCTGAAAACTTCCTCCCCAGGGCTCGGACAGGGCCCCTGTTTGCCTGTCTC 2040

Qy   2053 attgggaagcagatgaaggctctgcgggacggtgactggttttggtgggagaacagccac 2112
          ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   2041 ATTGGGAAGCAGATGAAGGCTCTGCGGGATGGTGACTGGTTTTGGTGGGAGAACAGCCAC 2100

Qy   2113 gtcttcacggatgcacagaggcgtgagctggagaagcactccctgtctcgggtcatctgt 2172
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GTCTTCACGGATGCACAGAGGCGTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGT 2160

Qy   2173 gacaacactggcctcaccagggtgcccatggatgccttccaagtcggcaaattccccgaa 2232
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   2161 GACAACACTGGCCTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCTGAA 2220

Qy   2233 gactttgagtcttgtgacagcatcactggcatgaacctggaggcctggagggaaaccttt 2292
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   2221 GACTTTGAGTCTTGTGACAGCATCCCTGGCATGAACCTGGAGGCCTGGAGGGAAACCTTT 2280

Qy   2293 cctcaagacgacaagtgtggcttcccagagagcgtggagaatggggactttgtgcactgt 2352
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 CCTCAAGACGACAAGTGTGGCTTCCCAGAGAGCGTGGAGAATGGGGACTTTGTGCACTGT 2340

Qy   2353 gaggagtctgggaggcgcgtgctggtgtattcctgccggcacgggtatgagctccaaggc 2412
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 GAGGAGTCTGGGAGGCGCGTGCTGGTGTATTCCTGCCGGCACGGGTATGAGCTCCAAGGC 2400

Qy   2413 cgggagcagctcacttgcacccaggaaggatgggatttccagcctcccctctgcaaagat 2472
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 CGGGAGCAGCTCACTTGCACCCAGGAAGGATGGGATTTCCAGCCTCCCCTCTGCAAAGAT 2460
```

Top alignment block (left column):

```
Qy  2473  gtgaacgagtctctgaacgcgtgccaccctgcagcgcctctggaggtgcagaaac  2532
          ||||||||||||||||||||| ||||||||||||||||| || |||||||||||||
Db  2461  gtgaacgagtctgcgaacgcgtgccacccttgcaccccttcgccaagtgcagaaac  2520

Qy  2533  accaaagcggctgcatgtctgtctctggagcaggaaccactgaggaacgacggagaa  2592
          || ||||| ||||||||||||||||||||||||||||||||| |||||||||||||
Db  2521  accaaagcgcgctgcatgtctcctccatgctgccctccgacgac gatggacga  2580

Qy  2593  acctgccagaactctcgaggcagggtgacctgcgccctgccttgcagtctgctgcctt  2652
          |||||||||||| |||||||||||||||  ||||||||||||||||||||||||||||
Db  2581  acctgccagaactctcgaggcagggtgaccttgcagatgctgcgctgcctt  2640

Qy  2653  ctgctatcctgcggagcctgtgacctcgcctgagctcgctggagcccgacct  2712
          || |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641  ctcatcctgcggagcctgtgacctcgcctgagctcgctggagcccgacacct  2700

Qy  2713  ggcactaaatccacactgcccatcctggaaccagcctgagctgcccctgtagggcc  2772
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701  gccactaaatccacactgcccatcctggaaccagcctgagctcccccgtgacatgc  2760

Qy  2773  ggaaagccaggccctgggactgagaacagctctgagactcatcctgcagggcgat  2832
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761  ggaaagccaggccctgggactgagaacagctctgagactcatcctgcagggcgat  2820

Qy  2833  gagatcgctgggatgaatactcacagctccctcgaagtgca  2892
          |||||||||||||||||||||||||||||||||||||||||
Db  2821  gagatcgctgggatgaatactcacagctccctcgaagtgca  2880

Qy  2893  acgttcgagaactcgaaagctcagtctccaaataccctttccca  2952
          |||||||||||||||||||||||||||||||||||||||||||||
Db  2881  acgttcgagaactcgaaagctcagtctccaaataccctttccca  2940

Qy  2953  agtggcagacactgcagacaggtgcatcgttagcaagtcagta  3012
          |||||||||||||||||||||||||||||||||||||||||||
Db  2941  aacacaggccaaaatcaagacactggaaaatgcgatacttgt  3000

Qy  3013  catgtcgagttactctcgatggatgaatagaatactgc  3060
          |||||||||||||||||||||||||||||||||||||
Db  3001  ccatctccagttactctcaggatgatgaatataatctgc  3048
```

Top alignment block (right column):

```
Qy    45  atattcagcatgcagttgctgctgagaaaagaagaaagatgacgccttgggctgtctgc  104
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  attactgcagtgcagttgttgctgcagagaaaagaaaaagaagatgacgccttgggctgtctgc  60

Qy   105  gtcgtcacgctcgtatggcctgccgagaagctctccctcatctcgagagggaa  164
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  gtcgtcacgctcgtatggcctgccgagaagctctccctcatctcgagagggaa  120

Qy   165  agaacctccttggccagccggtgggagtttggcatctcagttctgaaaaggaaga  224
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  agaacttccttggccagccggtgggagtttggcatctcagttctgaaaaggaaga  180

Qy   225  gcgcctgggaaaaccgtagccagccatctccgtcggcggagatatgagccaaccaa  284
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  gcgcctgggtgagaaccccatgtacccacgaacggagagaaggcaagagaaaggaa  240

Qy   285  cctttctggagcagctctgtctcttttccaaactcctctggccaaccaagccggtcat  344
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  cctttctggagcagctctgtctcttttccaaactcctctggccaaccaagccgggtgat  300

Qy   345  tgcccgacgcagagatagtagaaacatcaatcaacagcgtgaaagaaagcaacct  404
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  tgcccgacgcagagatagtagaaacatcaatcaacagcgtgaaagaaagcaacct  360

Qy   405  gaaaaccaacaaccagcacccaccacgagctccagaatctctcgcatcct  464
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  gaaaactcaacaatccagcatcatccgtttatacagaaagagctctgcaccat  420

Qy   465  tgcaaacatgtctggatgtctccctacgtgccccaaatagcccaaccctgcct  524
```

Right-hand feature/annotation listing:

```
protein_id="CAA59467.1"
/db_xref="GI:3725"
/db_xref="SWISS-PROT:P07202"
/translation="MRALAVLSYTLVMACTEAFFPFISRGRILLWGKPEQSRVSSVLR
SRLVDTYAMVMQMNQKLKRGISSPRAGLLSPRAGLLSPRLPKSPSWLANLAEMSTQAMK
RKVNLVKQGSQHFDALESDLSILNHLQPKLPKCPWTCLANKYRTCACNMNRALAVLSYTL
FDDPGRGSNTALARHLPVTFDDFSQPRGRNFDFLMGTPLLPVRERYRYIQVSNE
VPDDDAGGGAPPEVDNLLSPRNLLLVPRSPNSLAWSRLTPGDPGTPTGPLLQLSVLR
EAPPAAGTACLPFYRRRSSAACCTGNQGRILLRRLSTAPRPQMNWLLRSTYVSSSP
ALRQLPRMTSAEDLLAVVLRLRDSDRATLPVPPPRRAPACAPEGIPGTRGPFLA
LNSTLHYALMADRRRPSPVPTGPLLPEETMHWMEFAAGLCTASNIYGLSLLEVNLRI
PSRQWHQTRTHIRUPAL"
```

```
sig_peptide    1..43
               /product="put. signal peptide (AA -14 to -1)"
mat_peptide    83..2839
               /product="put. mature thyroperoxidase"
misc_feature   425..433
               /note="put. N-glycosylation site"
misc_feature   959..967
               /note="put. N-glycosylation site"
misc_feature   1064..1072
               /note="put. N-glycosylation site"
misc_feature   1472..1480
               /note="put. N-glycosylation site"
misc_feature   1471..2643
               /note="put. N-glycosylation site"
misc_feature   2579..2652
               /note="put. transmembrane region"
BASE COUNT     664 a     928 c     871 g     564 t
ORIGIN
```

```
Query Match  97.4%;  Score 2991.4;  DB 9;  Length 3027;
Best Local Similarity  97.4%;
Matches 3001;  Conservative 0;  Mismatches 16;  Indels 0;  Gaps 0;
```

Qy   2685 gacgtgtttcgaccggatgcagcctcgtcagcctctgctctttttccaaactctcctgaccaacacaagcga 2744
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 GACGGTGATTTCCAAGTGGACAGCACTGGCACTAAATCACACTGCCCAVTCTCGGCAGAC CAACAAGCGGA 2700

Qy   2745 agccggaggaactcccacgtcgtgatgcgacgagagcgttaggacctcacga 2804
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 AGGCGGAGGAACTCCCACAGCTGAGATCACAGAGAGAACACCAGCAGACAAGCAC 2760

Qy   2805 gcgggccctcagcctctcgagggtctcgaggtgggggttgagaagcggagactgactca 2864
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 GCGGGCCCGCACGTCAGGCACTCGACACGAGAGATGCTGGGATGGAAGGCCGGGATATACTA 2820

Qy   2865 caagctcagcgaaagcccctcgtgagagcaaatgtcgagacactgcagaacgactcatgtt 2924
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 CAGGCTCGCCACGAGACCCCTCTCAGGCCAAAGTGGCGACAGAGCAGCGTCATGTT 2880

Qy   2925 cccaaatcacgtacgactctttccaacagcggaatcgagcagcaccccaga 2984
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 CCCAAAATCACGTACGACTCTTTTCCAAACACGACCAATCAGCAGAACCACCCCAGA 2940

Qy   2985 tgtttcccaacacggtaaactagtacatgtcggatctatgagagcagctc 3044
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 TGTTTTCCCAACACGGGTAAACTGTAGTTACTCACGGCATGGATGAA 3000

Qy   3045 taaagtatagctgca 3061
           |||||||||||||||
Db   3001 TAAAATGTATAGCTGCA 3017

Query Match   91.2%;   Score 2803;   DB 6;   Length 2847;
Best Local Similarity 99.3%;   Pred. NO. 0;
Matches 2827;   Conservative 0;   Mismatches 15; Indels 5; Gaps 1;

Qy   45 attacctcagacatgcagtttggcctgaagaaggaagaaagcgctcgct 99
           ||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   1 ATTACCTCAGACATGCAGTTTTGGCCTGAAGAAGAGGGAAAAAAGCGCAGATAGAGCGCTCCT 60

Qy   100 gtcgtctcacgacctggagagactgagccaagaagagatcttctgccctcctccccatccctccggag 159
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GTGCTGTCTGTCACGTGCTGGAGAGGCTGAGCCAAGAGGAGATCTTCCTCCCTCATCTCCGAGA 120

Qy   160 gggaaagaacctcttgggagaagctgagggatcgtgtctctcaacgttggaagga 219
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGGAAAGAACCTCTTTGGGGAGAAGCTGAGGGATCGTGTCTCTCAACGTTGGAGGA 180

Qy   220 agcgaagcgctgtggacgcacgcacgacgacaagaaaagacgggag 279
           ||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AGCAAGCGCTGGTGGACAGCACAGTACGCACAGGAAAACTCAAGAAAGA 240

(This page presents a pairwise nucleotide sequence alignment — labeled "Qy" (query) and "Db" (database) — running from approximately position 1360 to 2859, followed by a GenBank-format sequence record. The dense aligned nucleotide lines are reproduced below in structured form.)

```
RESULT  6

LOCUS       HUMTPOB      2877 bp    mRNA            PRI       14-JAN-1995
DEFINITION  Human thyroid peroxidase mRNA, clone pHTPO-2.4.
ACCESSION   J02970
VERSION     J02970.1  GI:339870
KEYWORDS    alternatively spliced; thyroid peroxidase.
SOURCE      human thyroid gland (from patients with Grave's disease). cDNA to
            mRNA, clone pHTPO-2.4.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2877)
  AUTHORS   Kimura,S., Kotani,T., McBride,O.W., Umeki,K., Hirai,K., Nakayama,T.
            and Ohtaki,S.
  TITLE     Human thyroid peroxidase: complete cDNA and protein sequence,
            chromosome mapping, and identification of two alternatively spliced
            mRNAs
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 84 (16), 5555-5559 (1987)
  MEDLINE   87289641
COMMENT     Computer-readable copy of sequence in [1] kindly provided by
            S.Kimura, 09-SEP-1987.
FEATURES             Location/Qualifiers
     source          1..2877
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="2pter-p24"
     mRNA            <1..2877
                     /note="TPO mRNA"
     gene            73..2703
                     /gene="TPO"
     CDS             73..2703
                     /gene="TPO"
                     /note="thyroid peroxidase"
                     /db_xref="GDB:G00-120-446"
                     /protein_id="AAA61216.1"
                     /translation="MRALAVLSVTLVNACTEAFFTISRGELLLWGKPEESRVSVVLE
                     ESKRLVDTAMTAYMQQNRLKKRGILSPAQLLSFSKLPEPTSVIARAAEIMETSIQAMK
                     RKVNLATQQSQHPTDALSEEDLSIANMSCLPTHLPKPCPWNCLANKYRPITGACNN
```

```
RDHPRRTASMRPLLARWLPYVEDGFSQFRGRHRRGFLMGFPLPPVREVYFRIVTQVSNE
VVPTGDRWISDLLAWAKQVIDRDIRFTQDSYTRNAAAYQGDADQQTCKNGNPFSPYQLFE
ERRPAAGTAQLPFYSEASAACGTUDGALGFSHLSTARPKGOWNCLAPSTYDTSAGSFY
ALERQLRNWTEAGCLLRVYINALRDSGRAYLPPVPPPRRPAACAPERPCIPCETRCPCLA
GIGAASIYPISIAVLGYTARRIRQHLAALAVLNRNHRRDATVDDAVTYGAVKQALSEIAF
RGPITAYGLFSSAAACGTDGALGFSHLSTARPKGOWNCLAPSTYDTSAGSF
QRHPDLGCLWHLHQKFFSHWTLLRGGYTNKWRECGLLPRLEFPADLSTAIASRSVADKIL
GLYRHSRHNVUOGLARMRENFAPSPYGATFPVRVTVPAVRTYSIRSNRHPWNSISRYPTDDA
MRCXAQAHRPIKGLADGDPVHYRLADMMVERTAMWHEFIVVFRCTVDWSSHFSSGPPGVQSTI
DKCGFPSVEERDSFVHCEEESGRRVVLVSCRRGTELQGREGLTCTGCRGWNFQPDFPLCKDV
NRCANGAGMLPHANSAAENNTRTKDCFCQOLADFPFHGEDDSRPCVQSLPARTMISWKSLA
SEQESAGMEGRUTHRLPRAL*
SEQESAGMEGRUTHRLPRAL*
```

```
BASE COUNT     632 a      883 c      829 g      533 t
ORIGIN  Chromosome 2p:ar-q11.
```

```
Query Match    86.3%;  Score 2652;  DB 9;  Length 2877;
Best Local Similarity  93.7%;  Pred. No. 0;
Matches 2857;  Conservative  0;  Mismatches  15;  Indels 176;  Gaps  2;
```

```
Qy     18 catttcagaagggttacagccgtgaaaattactcagacgtcgagaagagga  77
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CANTTCAGAAGGGTTACAGCCGTCGAAAATTACTCAGACGTCGAGAAGAGGA  60

Qy     78 aaaa-------agaatgagccgtcgcgtcgtcgtatggccgtatcgccc  132
          |||||       |||||||||||||||||||||||||||||||||||||
Db     61 AAAAAGGTCAGAATGAGCCGGAGCCTCGTCGCGCTCGTCGTATGGCCGTCACA  120

Qy    133 gaagcctttccctcatctcgagagagaaccggggcctgtcctgccgagagagacgccg  192
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GAAGCCTTTCCCTCATCTCGAGAGGAAGAACCTCTTGGGACAAGCCTTGAGGAG  180

Qy    193 tctcgtttctcctagcgcttcgagagaagaatccttctggagctcagctccgtctctt  252
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCTCGTGTCTCCTAGCGCTTCGAGAGGAGGAAGGACCTCTGGGACAAGCCTCA  240

Qy    253 acgacagagagaaccagaaaggaatccctcaggaagtctctgacgcttgttctt  312
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACGATCGAGAAAACCTCAAGAAAAGGATCCTTCTGCAGCTCGTCTGTCTTTT  300

Qy    313 tccaaatcctgtgtctgccaacaggggatgcttggggctactatgtgaaaca  372
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCCAAATTCCTGAGCCAACAACCTGGGACCCGAGCACGCAGATAATGACAAACA  360

Qy    373 tcaatacaagcgtgaaaagtaaagtcaacctgaaatccaacaatcacagtatccccacg  432
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCANTACAACAGCGATGAAAAGAAAGTCAACCTCAAACATCCACAATCACAGTATCCCCACG  420

Qy    433 gatgcttatcagagctcggagctcagatgcagctcgcaaaactcggaccgcgctccttac  492
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GATGCTTATACAGAGATCTGCAGATCATGCAGCTGCAAAACTCGGACCTGCCTCCTTAC  480

Qy    493 atgctcgcccaaaatgtcctggacaacaggacgggcctccaacaggccatcaca  552
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATGCTCGCCCAAAATGTCCTGGACAACAATACAGGCCTCCAACAGGCCATCACAGA  540

Qy    553 gcttgacaacaggacccccaggatgggagcccttccaacagcggcgatgcgtatg  612
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GCTTGACAACAGGACACCCAGGATGGGAGCCCTTCCAACAGCGGCGGTGCGTATGG  600

Qy    613 ccccctccaastcagtcgaaaccctcgctgtcacaggggcaggtgcgctccgtcttcttg  672
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CCCCTCCAAGTCAGTCAAACCCTCACTGTCACAGGGGCAGGTGCGCTCCGTCTTCTTG  660

Qy    673 tacatagggttccccatcgacccgtgacataggcattgagaagctcattcaagtttca  732
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TACACAGGGTTCCCCATCGACCCGTGACATGGGCATTGAGAAGCTCATTCAAGTTTCA  720

Qy    733 aatgagggtttgtcagcagatagacccgtctcctgatgatggagcaaatac  792
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AATGAGGGTTTGTCAGCAGATAGACCCGTCTCGATGGCGATGGAGCAAATAC  780

Qy    793 atcgaacacacgatccgtgttcaacacccgagaccacccaaagccagcctgcttcttcgggagg  852
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ATTGAACACAGCATCCGTGTTCAACACACGAGACCACCCAAAGCCAGCCTGCTTCTTCGGGAGG  840

Qy    853 tcggatcccgatgaacttctggaacaaaacaccgatgttttccacaataacctccggaag  912
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GCTGACTCCCAAGGACTTGTGGACAAAACAACAGCAAAACCAATGTTTCCTTCCCATACAACTCCCGGAG  900

Qy    913 gagcgccgcgagggccgggagcagctgtcatgcctcaccgctcttccggctggctgctgc  972
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GAGCGCCGGCGGAGGGGCGGGAGCAGCTGTCTCCGGTCACCGCTTCGGCTGGCTGCTGC  960

Qy    973 ggcaccggaccaagcgcgtctttggagaccgcagcgactcagcagcatacgacgag  1032
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GGCACCGGGGACCAAGCGCGTCTTTGGGAGACGCGCAGCGACTCAGCAGGCATACGACGAG  1020

Qy   1033 atgaaggttgaacctgttcccgaggtccaaccgtatgggacgtcccccggcccta  1092
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ATGAACGGTTGACCTGTTCCCTGGACGGCGTCCCATCACCGGTATGGCAAGCTCCACCGGCCCTA  1080

Qy   1093 gaggcgagctcgcggaactgctgtcggcaggggtcaccggcccgccagccgcgaccc  1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GAGAGGCAGCTCGCGGAACTGCTGGCACAGTGCCGGAAGGGTCGTCGCACGCCGCCGGCGCGAG  1140

Qy   1141 cggaatcccggccgcgctactgcctcgtatggccgtcgcgctcgtcgtcggcgtgttgtcgg  1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CGGAATCCCGCGCGCTACTGCCTCGTATGGGCCGTCGCGCTCGTCGTCGGCGTGTTGTCGG  1200

Qy   1201 ccggagcccggaaccccggctcgtcctcagggagcacggctcctgtcgcggagagagcg  1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CCGGAGCCCGGAACCCCGGCTCGTCCTCAGGGAGCACGGCTCCTGTCGCGGAGAGAGCG  1260

Qy   1273 gccagagggtcccctgccccaagccgctgtagtcgcctcaagaagccgtgacaccc  1332
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GCCAGGAGGGTCCCTGCCCCAAGCCGCTGGACCGGCTCAAGCAGTGGCTGGCGGACAACACCC  1320

Qy   1333 ctgccgcggcctcaaggccctccaagctgaggcggagcggcgacggcgtatacaggag  1392
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTGCCGCGGCCTCAAGGCCCTCCAAGCTGAGGCGGAGCGGCGACGGCGTATACAGGAG  1380

Qy   1393 gctcgtcaaggctcttcagaagcatcaccsaggatatgtgaaaca-tacatcccagag  1452
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GCTCGCAAGGTCGTGGGCCTTCTCAGAGCATCATCACCAGAGGGAGTATGTGAAAACAACCC  1440

Qy   1453 actcagatgtatggactcaccatcaacctcacgaccatacaccacag  1512
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ATCCTGGACCTGGAACCGAGGCCTTCCACGAGGCTCTATGAAGGGTATGACTCCACC  1500

Qy   1513 gcaacaaccactgtgtcaagctgctggatgttcttcagccgctctcggccagccagc  1572
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCCAACACCAGTGTCCAAGCTGCTGGATGTGCTCAGCCGCTCCTGTTCGCCATCGCACG  1560

Qy   1573 accccccgcggccgggctcaaggagctcattcaggatgacccactccggatcgcgcggg  1632
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ACCCCGCGCGGCCCGGGCTCAAGGAGCTGATCCAGGACGACCTGCCGCGGATTGCCCGGG  1620

Qy   1633 ctgtgatctgacsacagacggcctccacaccagcggagctattcctggagtgtgttctggac  1692
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CTGTGCGTGCGCGACCAGCGTTTCTTTCTACGCCCATCGACACCATCACTCCGGTG-  1667

Qy   1693 ccactaatacgaggcttcttgcagagcagcaaaatcgaggtcaggtgcgagacagctcagatg  1752
          ||||||||||||||||||||||||||||||||||||||||||||
Db   1668 ----------------------------------------------  1667

Qy   1753 aacggagagtgcacggaagagtccttcttgtctgtccattccaagcaacttggactcggccgtg  1812
          ||||||||||||||||||||||||||||||||||||||||||||
Db   1668 -------------------------------------------------  1667

Qy   1813 tccatcaacctcgagtatgaccgtcattcaccacgagcggtcgccggttcacataatagggagggagg  1872
          ||||||||||||||||||||||||||||||||||||||||||||
Db   1813 ----------------------------AGGTTACAATCAGTCGACGGAGGGAG  1689

Qy   1873 tctctgcggcctgcctcggtcagaggagaccctcggtgactcggagcacgccatcgcagccagcagg  1932
          ||||||||||||||||||||||||||||||||||||||||||||
Db   1873 TACAACGGTTGCACAGATGACCCTATTCTGACCTCCCAGTTCAGG  1932
```

```
db  1690  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  1749
Qy  1933  agcgtggcgacttgctccgctcggacctctgaccanaaatcngatcctcgggctg

db  1810  GtggGcCTTAgCTGAAAACTTCTCTCCCCAGGCTCGACAGGAGCCCGGGTGTTTCCTGTCTC  1869
Qy  2053  attggaacagaacgaaggcctgcgactgttttgtggagagcagccac  2112
          |||||||||||||||||||||||||||||||||||||||||||||
db  1870  ATTGGGAACGAGATGAAGGCTCTCGAGGATGGTGACAGCACTGCCTCCGTGGCGGAACACCCAC  1929

db  1990  gacaacactggctcacacagggtgaccatcagatagcctccaaggtcggccaaaatccccaa  2049
Qy  2173  gacaacactggctcacacagggtgaccatcagatagcctccaaggt  2232

db  2050  gactttgatctctgtacagcctacacgggacgtcgggagagacatcct  2109
Qy  2233  gactttgatctctgtacagcctacacgggacgtcgggagagacatcct  2292

db  2170  gaggagagccgacaggtgcatgagaaggacatcggctactgt  2229
Qy  2353  gaggagagccgacaggtgcatgagaaggacatcggctactgt  2412

db  2290  ccccagcaagctctgtctcccgagaggtgaggaaaac  2349
Qy  2473  gtgaacaagtctgcaprggctgcgccaccgagctctg  2532

db  2350  accaaagggttccagtgttcaggtctcaggttaggagaacggaagaa  2409
Qy  2533  accaaagggttccagtgttcaggtctcaggttaggagaacggaagaa  2592

db  2410  accttcctaga actctgggaggcctccggagccgccggacc  2469
Qy  2653  cctgcggagccggagccggagcctggggcgggacgat  2712

db  2470  gaact ... gatgcagacaggcagggacggag  2529
Qy  2713  gaact ... gatgcagacaggcagggacggag  2772

db  2530  gaggagctcgatcggagccgggaccgcggcgctgccc  2589
Qy  2773  GGAAAGCATACCACTGCCCCCCGTGCAGGACC  2832

db  2590  ggagagctcgatcggatggaagccgggaccgctacacagactggctg  2649
Qy  2833  gaagtgtcggatcggatggaagccgggaccgctacacagactggctg  2892

db  2710  agtgcgagacctccagaacgggtccatgctgtccagctctagg  2769
Qy  2893  agtgcgagaccacccagaacgggtccatgctgtccagctctagg  2952

db  2770  AACACGGCAAATCGGACAAATCAGCACGACGACTGTTTTCCCAACACGGGTAAATCTAGTA  2829
Qy  2953  aacacggcaaatcggacaaatcagcacgacgactgttttcccaacacgggtaaatctagta  3012
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  3013  ccatgtcgaattttactctcggscatggtgaataaagtttatagctgc  3060

db  2830  CCATGTCTAACTTCTCTCCAGANGCATGGTGAATANATGATTTATANTGCTTACTGCTA  2877
```

RESULT 7
HSASTPO2
LOCUS   HSASTPO2   2845 bp   mRNA   PRI   12-SEP-1993
DEFINITION Human mRNA for alternatively spliced TPO 2 gene for thyroid
            peroxidase (EC 1.11.1.7).
ACCESSION   X17359
VERSION   X17359.1 GI:28897
KEYWORDS   autoantigen; peroxidase; thyroid peroxidase; TPO-2 gene;
            transmembrane protein.
SOURCE   Homo sapiens
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2845)
  AUTHORS   Barnett,P.S.
  JOURNAL   Submitted (20-DEC-1989) Barnett P.S., King's College School of
            Medicine & Dentistry, Denmark Hill, London SE5 8RX, United Kingdom
REFERENCE   2 (bases 1 to 2845)
  AUTHORS   Barnett,P.S., Banga,J.P., Watkins,J., Huang,G.C., Gluckman,D.R.,
            Page,M.J. and McGregor,A.M.
  TITLE   Nucleotide sequence of the alternatively spliced human thyroid
            peroxidase cDNA, TPO-2
  JOURNAL   Nucleic Acids Res. 18 (3), 670 (1990)
  MEDLINE   90175008
FEATURES   Location/Qualifiers
     source   1..2845
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2pter-p12"
                     /tissue_type="thyroid"
                     /clone_lib="lambda gt10"
                     /clone="TPO-21"
     CDS   66..596
                     /note="thyroid peroxidase (AA 1-876)"
                     /codon_start=1
                     /db_xref="GOA:P07202"
                     /db_xref="SWISS-PROT:P07202"
                     /db_xref="GI:28939"
                     /translation="MRALAVLSVLLWACTEAFFPFISRGKELKLWGKPEESRVSSVLE
                     ...
BASE COUNT   622 a   880 c   819 g   524 t
ORIGIN

Query Match   86.2%;
Best Local Similarity   94.0%;
Matches 2836; Mismatches   0; Conservative   0; Indels 171; Gaps   1;

```
Qy  20  ttcagagagctcacagccagcaatcagaaatcacacgcatcgataacgagagagagaa  79
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
db   1  aacgcaggcaaatcggacaaatcagcacgacgactgttttcccaacacgggtaaatctagta
```

```
Db      1 TTTCAGAAGAGTTACAGCCGTGAAAATTACTCAGCAGTGCAGTTGGCTGAGAAGAGGAAA 60

Qy     80 aaagaatgagagcgctggctgtgctgtctgtcacgctggttatggcctgcacagaagcct 139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAAGAATGAGAGCGCTGGCTGTGCTGTCTGTCACGCTGGTTATGGCCTGCACAGAAGCCT 120

Qy    140 tcttccccttcatctcgagagggaaagaactcctttggggaaagcctgaggagtctcgtg 199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGCCTGAGGAGTCTCGTG 180

Qy    200 tctctagcgtcttggaggaaagcaagcgcctggtggacaccgccatgtacgccacgatgc 259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCTCTAGCGTCTTGGAGGAAAGCAAGCGCCTGGTGGACACCGCCATGTACGCCACGATGC 240

Qy    260 agagaaacctcaagaaaagaggaatcctttctggagctcagcttctgtctttttccaaac 319
          ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db    241 AGAGAAACCTCAAGAAAAGAGGAATCCTTTCTCCAGCTCAGCTTCTGTCTTTTTCCAAAC 300

Qy    320 ttcctgagccaacaagcggagtgattgcccgagcagcagagataatggaaacatcaatac 379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTCCTGAGCCAACAAGCGGAGTGATTGCCCGAGCAGCAGAGATAATGGAAACATCAATAC 360

Qy    380 aagcgatgaaaagaaaagtcaacctgaaaactcaacaatcacagcatccaacggatgctt 439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAGCGATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGGATGCTT 420

Qy    440 tatcagaagatctgctgagcatcattgcaaacatgtctggatgtctcccttacatgctgc 499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTCTCCCTTACATGCTGC 480

Qy    500 ccccaaaatgcccaaacacttgcctggcgaacaaatacaggcccatcacaggagcttgca 559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CCCCAAAATGCCCAAACACTTGCCTGGCGAACAAATACAGGCCCATCACAGGAGCTTGCA 540

Qy    560 acaacagagaccaccccagatggggcgcctccaacacggccctggcacgatggctccctc 619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ACAACAGAGACCACCCCAGATGGGGCGCCTCCAACACGGCCCTGGCACGATGGCTCCCTC 600

Qy    620 cagtctatgaggacggcttcagtcagccccgaggctggaaccccggcttcttgtacaacg 679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CAGTCTATGAGGACGGCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTGTACAACG 660

Qy    680 ggttcccactgcccccggtccgggaggtgacaagacatgtcattcaagtttcaaatgagg 739
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GGTTCCCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAATGAGG 720

Qy    740 ttgtcacagatgatgaccgctattctgacctcctgatggcatggggacaatacatcgacc 799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGGGACAATACATCGACC 780

Qy    800 acgacatcgcgttcacaccacagagcaccagcaaagctgccttcgggggagggtctgact 859
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    781 ACGACATCGCGTTCACACCACAGAGCACCAGCAAAGCTGCCTTCGGGGGAGGGGCTGACT 840

Qy    860 gccagatgacttgtgagaaccaaaacccatgttttcccatacaactcccggaggaggccc 919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GCCAGATGACTTGTGAGAACCAAAACCCATGTTTTCCCATACAACTCCCGGAGGAGGCCC 900

Qy    920 ggccggccgcgggcaccgcctgtctgcccttctaccgctcttcggccgcctgcggcaccg 979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GGCCGGCCGCGGGCACCGCCTGTCTGCCCTTCTACCGCTCTTCGGCCGCCTGCGGCACCG 960

Qy    980 gggaccaaggcgcgctctttgggaacctgtccacggccaacccgaggcagcagatgaacg 1039
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db    961 GGGACCAAGGCGCGCTCTTTGGGAACCTGTCCACGGCCAACCCGCGGCAGCAGATGAACG 1020

Qy   1040 ggttgacctcgttcctggacgcgtccaccgtgtatggcagctccccggccctagagaggc 1099
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GGTTGACCTCGTTCCTGGACGCGTCCACCGTGTATGGCAGCTCCCCGGCCCTAGAGAGGC 1080

Qy   1100 agctgcggaactggaccagtgccgaagggctgctccgcgtccacggccgcctccgggact 1159
          |||||||||||||||||||||||||||||||||||||||||||||  |||||||||||||
Db   1081 AGCTGCGGAACTGGACCAGTGCCGAAGGGCTGCTCCGCGTCCACGCGCGCCTCCGGGACT 1140

Qy   1160 ccggccgcgcctacctgcccttcgtgccgccacgcgcgcctgcggcctgtgcgcccgagc 1219
          ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1141 CCGGCCGCGCCTACCTGCCCTTCGTGCCGCCACGCGCGCCTTCGGCCTGTGCGCCCGAGC 1200

Qy   1220 ccggcaaccccggagagacccgcgggccctgcttcctggccggagacggccgcgccagcg 1279
          |||||| |||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   1201 CCGGCATCCCCGGAGAGACCCGCGGGCCCTGCTTCCTGGCCGGAGACGGCCGCGCCACCG 1260

Qy   1280 aggtcccctccctgacggcactgcacacgctgtggctgcgcgagcacaaccgcctggccg 1339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AGGTCCCCTCCCTGACGGCACTGCACACGCTGTGGCTGCGCGAGCACAACCGCCTGGCCG 1320

Qy   1340 cggcgctcaaggccctcaatgcgcactggagcgcggacgccgtgtaccaggaggcgcgca 1399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CGGCGCTCAAGGCCCTCAATGCGCACTGGAGCGCGGACGCCGTGTACCAGGAGGCGCGCA 1380

Qy   1400 aggtcgtgggcgctctgcaccagatcatcaccctgagggattacatccccaggatcctgg 1459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AGGTCGTGGGCGCTCTGCACCAGATCATCACCCTGAGGGATTACATCCCCAGGATCCTGG 1440

Qy   1460 gacccgaggccttccagcagtacgtgggtccctatgaaggctatgactccaccgccaacc 1519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GACCCGAGGCCTTCCAGCAGTACGTGGGTCCCTATGAAGGCTATGACTCCACCGCCAACC 1500

Qy   1520 ccactgtgtccaacgtgttctccacagccgccttccgcttcggccatgccacgatccacc 1579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CCACTGTGTCCAACGTGTTCTCCACAGCCGCCTTCCGCTTCGGCCATGCCACGATCCACC 1560

Qy   1580 cgctggtgaggaggctggacgccagcttccaggagcaccccgacctgcccgggctgtggc 1639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CGCTGGTGAGGAGGCTGGACGCCAGCTTCCAGGAGCACCCCGACCTGCCCGGGCTGTGGC 1620

Qy   1640 tgcaccaggctttcttcagcccatggacattactccgtggaggtggtttggacccactaa 1699
          ||||||||||||||||||||||||||||||||||||||||
Db   1621 TGCACCAGGCTTTCTTCAGCCCATGGACATTACTCCGTGG-------------------- 1660

Qy   1700 tacgaggccttcttgcaagaccagccaaactgcaggtgcaggatcagctgatgaacgagg 1759

Db   1661 ------------------------------------------------------------ 1660

Qy   1760 agctgacggaaaggctctttgtgctgtccaattccagcaccttggatctggcgtccatca 1819

Db   1661 ------------------------------------------------------------ 1660

Qy   1820 acctgcagaggggccgggaccacgggctgccaggttacaatgagtggagggagttctgcg 1879
                                         |||||||||||||||||||||||||||||
Db   1661 -------------------------------AGGTTACAATGAGTGGAGGGAGTTCTGCG 1689

Qy   1880 gcctgcctcgcctggagacccccgctgacctgagcacagccatcgccagcaggagcgtgg 1939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1690 GCCTGCCTCGCCTGGAGACCCCCGCTGACCTGAGCACAGCCATCGCCAGCAGGAGCGTGG 1749

Qy   1940 ccgacaagatcctggacttgtacaagcatcctgacaacatcgatgtctggctgggaggct 1999
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1750 CCGACAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATGTCTGGCTGGGAGGCT 1809

Qy   2000 tagctgaaaacttcctccccagggctcggacagggcccctgtttgcctgtctcattggga 2059
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1810 TAGCTGAAAACTTCCTCCCCAGGGCTCGGACAGGGCCCCTGTTTGCCTGTCTCATTGGGA 1869

Qy   2060 agcagatgaaggctctgcgggacggtgactggttttggtgggagaacagccacgtcttca 2119
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1870 AGCAGATGAAGGCTCTGCGGGACGGTGACTGGTTTTGGTGGGAGAACAGCCACGTCTTCA 1929

Qy   2120 cggatgcacagaggcgtgagctggagaagcactccctgtctcgggtcatctgtgacaaca 2179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1930 CGGATGCACAGAGGCGTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACA 1989

Qy   2180 ctggcctcaccagggtgcccatggatgccttccaagtcggcaaattccccgaagactttg 2239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1990 CTGGCCTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCCGAAGACTTTG 2049
```

COMMENT

```
SRL_NC
OS   Unidentified
PN   JP 199904833-A/1
PD   09-APR-1999
PF   19-SEP-1997 JP 1997273743
PI   MASNO FUKUSHIMA
PC   C12N15/53,C12N9/08,C12N15/09,G01N33/531,G01N33/564//(C12N9/08,
PC   C12R1:91),
PC   C12N15/00
PC   Standedness: Double;
PC   Topology: Linear;
FH   Key              Location/Qualifiers
FT   source           /organism="Unidentified"
FT                    /mol_type=
FT                    /organism="unidentified"
FT                    /db_xref="taxon:32644"
```

FEATURES   Location/Qualifiers
  source   2546

BASE COUNT 542 a   804 c   725 g   475 t
ORIGIN

Query Match   81.8%; Score 2511.4; DB 6; Length 2546;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2524; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
Qy     85 atgasagcsgctsgctgctgtcscacgctgtatsgctgcacagaagsctttc 144
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAGAGCGTGCTSGTGGCTGTCSCACGCTGGTATSGCTSCACAGAAGSGCTTCTTC 60

Qy    145 ccttcatctcggagasasgaactccttsggsaagctsgsgsattcgtscgtct 204
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCCTTCATCTCGCACGCGAAAGACTCCTTYGGGGAACCTCGAGCAGTCTCGTGTCTCT 120

Qy    205 agctctctcggaggcagccgcctggtggacaccgccatstacgccgacgtgcagga 264
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGCCTCTTCGGAGGAAAAACAGCCGCCTGGTGGACACCGCCATSTACGCCGACGTACAGGA 180

Qy    265 aacctcaagaaagagagagaatcctcctcggagccgcctctcttctcccaaacctcct 324
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AACCTCAAGAAAGAGAGAGAATCCTCTTCCGAGCCGCACTTCTCTTCTCCCAAACCTCCT 240

Qy    325 gagccaacacasgsgtgattsgccgsgagsasgttsatsgaaactcaatacagagg 384
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGCCAACAACGGATGAGTTGCCCAGCGACGAGATATTGAAACATCAATACAGAGG 300

Qy    385 atgaaaagaaacttctggagcctcaacaatcacactctgsagctcccatgtc 444
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ATGAAAAGAAACTTCTGGAGCCTCAACAATCACACTCTGGAGCTCCCCATGTC 360

Qy    445 gaagacctctgtagcaatatctcaaacatcctctggatgtccctacctgcccca 504
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GAAGATCTCTGTAGCAATATCTCAAACATCTCCGGATGTCCCTACCTGCCCCA 420

Qy    505 aaatcccaaacactctgcggcggagsssaaaaasagcctccctcaagagtctsgsaacaac 564
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AAATGCCCAAACACTCTGCGGCGGAGAAAAAAAAAGCCCCATCACAGAGCCTGCAACAAC 480

Qy    565 aagcaccccaactgagtgggsgcctccaactsccctccacscctsccaac 624
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGAGACACCCCAAGTGGGGCGCCTCCAACAGCGGCCTCCCCTTCCAGTC 540

Qy    625 tatgagacgcctcacgctcggcccgsgactgsaccccgsctcctgtacacsgggtcc 684
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TATGAGACGCCTCACGGCTCGGCCCGGGACTGGACCCCGGCTCCTGTACACGGGTCC 600

Qy    685 ccastgscccccsgtccggstgscaasgagstscaatcsaatsgsagttstsc 744
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CCACTGCCCCCCGGTCCGGGGTGACAAGAGGTCATTCAAATGTCAAGTAGGGTTGTC 660

Qy    745 acagatgatgaccgcctattcctgacctcctgtagctgcagcgaccacgacgc 804
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    745 ACAGATGATGACCGCTATTCCTGACCTCCTGTAGCTGCAGCGACCACGACGC 804
```

```
Qy   2240 agtcttgsgagsgsctscgsgctsgsgcsgssgcsgaggsaaccttctcaag 2299
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2050 AGTCTTGTGAGCAGCGATGCTGCATGAGACACTGGAGGCAAACCTTTCCTCAAG 2109

Qy   2300 acgacagtgtggctcccagsgsactgsasasgssgsctttgtgscactsgtscgagt 2359
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2110 ACGACAAGTGTGGCTTCCCAGAGACGTGGAGAATGGCGACTTTGTGCACTGTGCGAGT 2169

Qy   2360 ctsgagsgssgcscsgtgstgscttsgscgtsgsgstccsgscssgsggsggc 2419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2170 CTGGAGAGCGGCGTCTGTGTATTCCTGCGCACGGTATGAACTCCCAAGGCCGGGAAC 2229

Qy   2420 agctcactgscsgsasgssgsgsscggstsgstcsgstcccssgsssgtgsacg 2479
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2230 AGCTCACTTGCACCGAGSGATGGGATTTCCASCCTCCCCTGCCAAGATGTGAACG 2289

Qy   2480 asgtgsgagsgstgsccascscccccsgscasgsgcctscgsgstgscagsaaacsasagsg 2539
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2290 AGTGTGCAGACGGTGCCACCCCACCCCCTGCCASGGTCCGCAGAAACACCAAAG 2349

Qy   2540 gcygctccscsgtgtctccgsgsgsgsgsgssacagsgtsgsgsgaascctgcs 2599
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2350 GCGGCTTCCACGTGTCTCCGGAGSGCCGACCCCTSACCAAGTTAGGGACAAAACCTGCCS 2409

Qy   2600 tsgsctcsgsggsgsssasstsgssattcsscsgststctcsgstsgstcscstccs 2659
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2410 TAGACTCCGGTGCCTCCTGGGGGCGTGAATTGGATCTCSCAGTGTCGCTSCTCTSTCTA 2469

Qy   2660 tcsgagsctsccsgsgstccacstcsgsgtsgtttsgsgsgssgsgsgsgsgcscsgscts 2719
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2470 TCGGAGSGCTTCCCAGSTCCACCTCGCACGGTTSATTTGCAGGTSGACACGACSTTSGSCTA 2529

Qy   2720 tatscscscstcssgststssgsgscgssgssgstttccsssgssgsgsgsgsgsgc 2779
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2530 AATCCACACTGCCGCCCATTGSGSGSGCAGSGSGGSAAGAACTSCCCASSGSGSTGsAGSGSSAAAC 2589

Qy   2780 acsagstctgsgssgstscscscssgsgssgscsstsgssgstsgsgsgssgstsgsgsg 2839
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2590 ACAGSCTSGSAGSGSAAACCAGSCSCSCTGSGGSGACACSGGSSGSGSTSSGSSGSCSCASGAGA 2649

Qy   2840 csggstsgssgstsgssgsaatcsgsgssgssgsgsssgsgssgsgsgsgsgsgsgca 2899
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2650 CTGGSATSGSAAGSGCCGGSATACTTCACAGSTSCSGSCSCAGSCSCASGSGSGSCSCAAASGSGCA 2709

Qy   2900 csgststsggsgsscsgssgstsasstscscsgssgsssgssgststssgsgssgsgsgsgsgsgcsss 2959
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2710 CTGGSATSGSAAGSGCCGGSATACTTCACAGSTSCSGSCSCAGSCSCASGSGSGSCSCASGSGSGSCSCAAASGSGCG 2769

Qy   2960 gsaastcsgsasatcsgsgsssgsasgstctttsscscsasascgsgstsasastsctascsstscsts 3019
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2770 GCAACTGCAGACAGCAGCAGACTCTCTTCCASCACGGGTAAAATTTACTACATGTC 2829

Qy   3020 gtsgstsacstccsgcs 3035
          ||||||||||||||
Db   2830 GTAGTTACTCTCAGCC 2845
```

```
RESULT        8
E23825
E23825
LOCUS       E23825      2546 bp    DNA                          PAT     07-FEB-2001
DEFINITION  Antigen for immunoassaying antihuman thyroid peroxidase antibody
            //anti recombinant human thyroid peroxidase
ACCESSION   E23825
VERSION     E23825.1  GI:13024570
KEYWORDS    JP 199904833-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified sequences.
REFERENCE   1  (bases 1 to 2546)
  AUTHORS   Masno,F.
  TITLE     Antigen for immunoassaying antihuman thyroid peroxidase antibody
            //anti recombinant human thyroid peroxidase
  JOURNAL   Patent: JP 199904833-A 1 09-APR-1999;
```

```
Db   661  ACAGATGATGACCGGCTATTCTGATGGCATGGGCACAATACAYCGACCGAGAC  720

Qy   805  atcggcttcacacacacgagacccggctggccgtgggagggagtgtctgactcgccag  864
Db   721  ATCGGCTTCACACACGAGACARCAACGGGTCCCTTGGCGGAGAGGCGGTGGCCGAC  780

Qy   865  atgactctgagaacccaaaaccaagtgtttccccataacaactccccggagagaccggagcg  924
Db   781  ATGACTCTGAGAACCAAAACACAATCTGACTTTCGGCTGGCAGCTTTAGCT  924

Qy   925  gcccgggagcccggcggcctcctttacgcttcgcctccgggaggaggaa  984
Db   841  GCCCGGGAGCCCGCGGCCTTTCGGCCCCTTGCGGCCGGGAGGAGGGTAC  900

Qy   985  caggcgacgactctgggaactctgggagccgaaccggccgggatcggaggtg  1044
Db   901  CAAGGCGACGACTCTTGGGAACTCTGGGAGGAGGACAGCCACTGTCTTCACGGAT  960

Qy  1045  acctcgtcctgacgcgacccgtatgatgaccggaccctagagagcgagctg  1104
Db   961  ACCTCGTCCTGACGCCGGGTGCTGACAGAACAACTCCGGTTAAAGACAACACTGGC  1020

Qy  1105  cgganctgaccacgggcagaggctgctcccggccgcttccgagcctttccag  1164
Db  1021  CGAACTGACCACTGGCCAGAGGCTGCTCCCAGCCGCCTTCCGAGACCCTCGGC  1080

Qy  1165  ccaaggcctccaatggcgacaggaccaggcccgcttacgccgtcccaaggt  1224
Db  1081  CCAAGGCCTCCAATGGGCGACAGGACCAGGCCCGCCTTACGCCGTCCCAAGGTG  1140

Qy  1225  aaccccggagaccccgggcccgctgctccgggccgagcggcccgcagtc  1284
Db  1141  AACCCCGGAGACCCCGGGCCCGCTGCTCCGGGCCGAGCGGCCCGCAGTC  1200

Qy  1285  ccctccccaagctggcagggcaagacccccgctggggcggccgacggt  1344
Db  1201  CCCTCCCCCAAGCTGGCAGGGCAACGACCCCCGCTGGGGCCGACGGGT  1260

Qy  1345  ccaaggccctcaatggcgactgagaccctccaggacctgggtgcaggtc  1404
Db  1261  CTCAAGGCCTCAATGGCGACTGAGCGGTGGACAGACTCAGGACCAAGGTC  1320

Qy  1405  gtggggcctccgcacaggatctcacccgggagggatctcgggaccc  1464
Db  1321  GTGGGGCCTCCGCACACAGGATCTCACCCGGGAGGGATCTCGGGACCC  1380

Qy  1465  gggggctcccccaagctacctggacggtcactgaaggccccacacca  1524
Db  1381  GGGGGCTCCCCCAAGCTCCTGGACGGTCACTGAAGGCCCCACACCA  1440

Qy  1525  gtgtccaagctgcggcaccgcaccccgggtcatggct  1584
Db  1441  GTGTCCAAGCTGCGGCACCGCACCCCGGGTCATGGCT  1500

Qy  1585  gtgagagaggctgagggcaaaggcggcctgttggccccggactgag  1644
Db  1501  GTGAGAGAGGCTGAGGGCAAAGGCGGCCTGTTGGCCCCGGACTGAG  1560

Qy  1645  ggccttcttccagccagcttgacgaacacacggtgttcccagactaatagca  1704
Db  1561  GGCCTTCTTCCAGCCAGCTTGACGAACACACGGTGTTCCCAGACTAATACGA  1620

Qy  1705  ggcttctcagagccagaccaatccagtcgcagatctaaccctggagg  1764
Db  1621  GGCTTCTCAGAGCCAGACCAATCCAGTCGCAGATCTAACCCTGGAGG  1680

Qy  1765  acggaaaggtcttctgctgtccaatccacctggactgcgccatcaactg  1824
Db  1681  ACGGAAAGGTCTTCTGCTGTCCAATCCACCTGGACTGCGCCATCAACTG  1740

Qy  1825  cacagaggcgcggacacggggcctcgcattacataggcggagtctcggact  1884
Db  1741  CACAGAGGCGCGGACACGGGGCCTCGCATTACATAGGCGGAGTCTCGGGCTG  1800
```

```
Qy  1885  cctcccctgggacccctgccgttagacgacctgagagacgcccgcagcag  1944
Db  1801  CTTCCCCTGGGACCCCTGCCGTTAGACGACCTCCAGCCGCACGAGACGGTGCCGAC  1860

Qy  1945  gagttacggcttgcccgatcctcgacaacatcacatgtctgctgggagaagttgct  2004
Db  1861  AAGATCCTGGGATACCATCCTGACAACATCACATCGATTTCGGCTGGGAGCTTAGCT  1920

Qy  2005  gaaaacttcctcccaaggctcgacagggcccctcgtttgctctgctccatggagacag  2064
Db  1921  GAAAACTTCCTCCCGGCTCGACAGGGCCCCGTCTGCTCCATTGGAGCAG  1980

Qy  2065  atgaagtcctccgggagggacctggcactgcacctttcacggt  2124
Db  1981  ATGAAGACTCTGGGGAGGACCTGGCACTGTTTGTGCGGAGAAGCAGCCACCTCTTCACGGAT  2040

Qy  2125  gcacgaaggcgtctgaactgacagaagcactccctgtcgggtcatccggtcaacaacctggc  2184
Db  2041  GCACGAAGGCGTGAGCTGACAGAAGCACTCCCTGTCGGGTCATCCGGTGACAACAACCTGGC  2100

Qy  2185  ctcacacaggttgcccagtggatgcctccaggcgacaaatcccagctcccaagctcgggctt  2244
Db  2101  CTCACACAGGTGCCCAGTGGATGCCTCCAGGCGACAAATCCCAGCTCCCAAGCTCGAGTCT  2160

Qy  2245  tctggagcatcggatgaaccgggccggaccgagagaaaccttcccaagaaccggt  2304
Db  2161  TGTGAGCAGCATCACTGACGCATGAACCTGAGCGGCTGGAGGAAACCTTCCCTCAAGAACGAC  2220

Qy  2305  gagtcggctgtgggctcggcggagaatggggactttggtcactcagacgagagctcgg  2364
Db  2221  GAGTCGGCTGTGGGGCTCGGCGGAGAATGGGGGACTTTGTGCACTGTGACGAGAGTCTCGG  2280

Qy  2365  aggcggtgctcgggtatgtcgcccgggagtatgagctcccaaggccgggagcgggagtc  2424
Db  2281  AGGCGGTGCTCGGGTATGTCTGCCCGGGAGTATGAGCTCCCAAGGCCGGGAGCGGGAGGTC  2340

Qy  2425  actccacccaagagggagattcccgtcccccctccctgaagggaaacgagctgagtcgt  2484
Db  2341  ACTCCACCCAAGAAGAGGATCCTGTCCCCCCTCCCTGAAGGAGTGGAAAGTG  2400

Qy  2485  gagaagggttcccctcacatggagtgtcgagaacaccaaaggacgcggc  2544
Db  2401  GCAGACGGTGCCCCTCACATGGAGTGTCGCACAGACACCAAAGGACGCGGC  2460

Qy  2545  tccagtgtcctcgggtgacccctacggaggttaaggagaacctgcgtagac  2604
Db  2461  TTCCAGTGTCCTCGGGTGACCCCTACGAGTTAAGGAGAACCTGCGTAGAC  2520

Qy  2605  tccggagcctccctcgggttgact  2629
Db  2521  TCCGGAGACCTCCCTCGGGCCACCT  2545
```

```
RESULT   9
RNPO
LOCUS       RNPO        3237 bp  mRNA            ROD       12-SEP-1993
DEFINITION  Rat mRNA for thyroid peroxidase.
ACCESSION   X17396 M27275
VERSION     X17396.1  GI:57382
KEYWORDS    thyroid peroxidase.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus
REFERENCE   1  (bases 1 to 3237)
  AUTHORS   Rapoport,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-1989) to the EMBL/GenBank/DDBJ databases
REFERENCE   2  (bases 1 to 3237)
  AUTHORS   Derwahl,M., Seto,P. and Rapoport,B.
  TITLE     Complete nucleotide sequence of the cDNA for thyroid peroxidase in
```

```
JOURNAL    FRTL5 rat thyroid cells
MEDLINE    Nucleic Acids Res. 17 (20), 8380 (1989)
           90045972
FEATURES             Location/Qualifiers
     source          1..3237
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /note="rat"="FRTL5."
     CDS             42..2786
                     /note="thyroid peroxidase"
                     /protein_id="CAA35257.1"
                     /db_xref="GI:57383"
                     /db_xref="PID:e1319"=PID:e1319*="SWISS-PROT:P14650"
                     /translation="MRTLGAMAYMLVVWGTAIFLPFLLASRDLGGKYMSHIVSISVVE
                     TSGLVVDNAYYSTMKEMLKKRGILSPAQLLSFSKLPEPTSKAISRAAEIMETSIQVMK
                     ...
                     VICHWTNAQKKSTLLIYTERVIMSSGFRRSQESSISPQKAKVQVQADEPAVGSRVLCLE"

BASE COUNT     816 a    874 c    831 g    716 t
ORIGIN
```

```
Query Match       53.9%;  Score 1655.2;  DB 10;  Length 3237;
Best Local Similarity   76.3%;  Pred. No. 1.4e-292;
Matches 2105;  Conservative    53;  Mismatches 613;  Indels  42;  Gaps  4;
```

```
Qy   2782  caggccgtaggagacctcaccgcagccgcccgacctccggagcgagaagactgct  2841
           ||||| ||||||||||||||||||||| |||||||| ||||| || ||||||||||
Db   2697  CAGGAGAGTGGGATTGTCACCACAAAAGGCCGAGGTTCAAGATGTTGAACAGAACCGCCT  2756
```

```
BASE COUNT    819 a    897 c    854 g    721 t
ORIGIN
Query Match       53.0%;   Score 1628.6;  DB 10;  Length 3291;
Best Local Similarity   75.7%;   Pred. No. 9.7e-288;
Matches 2088;   Mismatches 629;  Conservative 0;  Indels  42;  Gaps   4;
```

```
Qy    82 agaataaagccgtggctcgtgctgtcaagcgttatggctcgcacagagctc 141
         ||||| | ||| ||  |||| || ||||| |  || ||| ||| |||| |||
Db    71 AGAATGAGAACACTTGGAGCTAATGCCGATCTGGTCGTTATGGGAACTGTATTTTC 130

Qy   142 ttccccctcatctcgagagggaaagaactccttggggaaagcctgagcctcgtgc 201
         |||  | ||| || ||| ||||| ||| || ||||||| || ||| |||| |||
Db   131 CTCTCTTTTATCCTGAGAAGCAGAGAACATCTTGTGGGAGAAGAGCCAATGAGTCCCCATGTT 190

Qy   191 ATCAGTGCTGTGGAAACCAAGCCAGCTCATGGTGGACCATGACAACACCATGAAA 250

Qy   202 tctagctgtctgggacgaagcgcgtgctgcgctcgatgctggc 261

Qy   251 AGAAACCTCAAGAAAAGGGAACGTCTTTCTGCACCCAGTCTGCTTTCTTTAAGCTG 310

Qy   262 agaaactccataaaagaaatctcttccgaacctccgctctgtctttctcaacct 321

Qy   311 CCCAAGTCCACCATCTGCAGCTGCACGAAGGATTAATGCCAACAACACATCAATACAA 370

Qy   322 cctgagccaacaacgagtcgatggcatggcagaataatggaacatccatacaa 381

Qy   371 GTCATGAAACG 381

Qy   382 gcgatgaaagaaagtccaacatcaccaatcaacatagctcaacggactgcttta 441

Qy   412 TGAACAGTCACAGTTCTCCACAGGATGCCTTA 412

Qy   442 tacgaaactctggatcatcgaaaacgctgagtgtctcctacatgtcgcct 501

Qy   413 TCACGAGCATTCTGCGACATTGTAACCACTGCTCCATGCACCAGTCCATCA 472

Qy   502 ccaaatgccaaaactctgctctgcgaacaaataagcgccatcacaggcgctgcaac 561

Qy   473 CCAAGATGTCTGACACTGCTGGCAAATAAGTACCGCCGCCCATCACAGGGGCGTGCAAC 532

Qy   562 aacgagacacccaatacggagagcgcctggagtggaatcgcttcaccacta 621

Qy   533 AATAGAAGATCACCCCAGATGGGGAGGATCCTCCAACACAGCCTCAAGATGCGCTGCCTCCT 592

Qy   622 gtcataagagaaccccgattttctgccggtgggacaacactacggtggc 681

Qy   593 GTCTATGAAGATGTTTCAAGTAGCCCAAAGGCTGGAACCCTAATTTCTATTACAGCGGC 652

Qy   682 ttccaaatgcccccggatggactgaaaaacatcctcaagttccaaatgaagttc 741

Qy   653 TTCACGAGTTCCTTGATACTCAAGCAGGCTGGACTACTAGAATCCAAGATGGCTGCCTTT 712

Qy   742 gtcacagtagagacatctcgacttcctcgatgtgatgggcatacatcgacac 801

Qy   713 GTGACGAAGATGACACTACTGTATTTTCTGCCGGTGTGACCGCTGACTGCTATTGCGCTTACT 772

Qy   802 gaatccgcttccaccacggacaccacaaacaagctgcctctgggggaggtgctcgactgc 861

Qy   773 GACATTGTCCACACACACAGACACAGAACAGTTACGAGGACATCCAGTCCAGGACAAGGCGGCAT 832

Qy   862 caaagtacttgagaaacaaaccccagttttcccataacctttcccccaaactcc 921

Qy   833 CAGCTGACCTGTAGACACCAAAATCTTTCTTTCTCCACTAACATCTTCCCAAACTCC 889

Qy   922 gtcacagttcttcctcccgggggatcgacaaggggccctcgacggagggaggaaa 981

Qy   890 ---TCAGGACACTGGCAATGCCTGATGCCTTTCACCAGTCCTTCTTGCATGTATCATTGGAGAG 946

Qy   982 gaccaaagcgcgctcttggggaacctctgaccccaaccggggagacgatgaaacgg 1041

Qy   947 GAGCAAGGCTGTCTTCCTGACATTGTCACTCCTCGGTCTCATCCTGGAGGAACAACATGTC 1006

Qy  1042 ttcaccctctctggacgctcaagctgaagctccccgtcctcctcagaaagcgag 1101

Qy  1007 TTGACTCCTTCTTGATCGTTCCACTGTACTCAGTCCTGAAGAAGCAG 1066

Qy  1102 ctcccgaatgaacctgccagggggccctcccggatcc 1161

Qy  1067 CTGGACAACAGATTACTGTCGGCAAGACTTTGGCCTTTGAAAGCTGAAGTTTGAA 1126

Qy  1162 ggccgcgctaactgccttcgtgcccgacgtgcgccttgtgccgagcc 1221
```

```
Db  1127 GGCCGTGCTACTTGCCTT--- 1174

Qy  1222 gcaaccccgggaaccccgcgg.................cccacacaggcccgcgccgtcgccagcag 1281
         ||  | |  || ||| || ||                 |  |  || |||||| | ||| ||||
Db  1175 GGTACCCAGCAACCACGACCGCCTGCTTCCTGCGTGGAGACGGTGCGCCTCCTGCTCAC 1234

Qy  1282 gtcccccccctaccgactacacccctatagtcggctgcaacctgctccgg 1341
Db  1235 GTCCCTGCCTGGCAGCGAGGTCACACACTTGTGGCCTGCGGAGCAACCGAGCCACCCACCTGGCCTTCG 1294

Qy  1342 ggctcaagccccaatgcgacgggctcctggaggaggtctaccagagcgtccgcaag 1401
Db  1295 GCCTTCAAGGCCATTAACAAGCACGTGGACCCAACATGCGATCACACACGCGGAGGAGCCCAG 1354

Qy  1402 gcgtggggtctgccagaatcaccagttactcatcctccagcagctggatactc 1461
Db  1355 GTGGTGAGGGCAGTGCCAGAGTATGTGGGCCCCATTGACGGCCTTAACACCTGGAAACCT 1414

Qy  1462 ccgagccctccacacgactactctgctcccagtacaccatccagcccaaacccc 1521
Db  1415 CCTGATGCCTTAGGGCAGTATGTGGGCCTCACTGTGGCTTTGGCCCATCCACACACAGTTGCAACCT 1474

Qy  1522 actgtggagaggccgagccaccgagtccaagaacccagacccaacctccggttcttcctg 1581
Db  1475 ACTGTGTCCAACATCTTCTCCAACTGTGACTTGGGCTCCTGATGCCCAGTGCAGCTG 1534

Qy  1582 tgcttggagggctacgagttcaggaagaaccccctggggtctgttcgacctata 1641
Db  1535 CTTGTTAGCAGCATGGAACTTACACTCTGGGCACTTGGGCCATGCAACAGTCCATCA 1594

Qy  1642 cacagcttcttcacccgactgacataccctgagttgtttgacagtcatacctcta 1701
Db  1595 CGTGATGTCTTCTTCAGACCTGCTGGCACAGTCAGCAAGGTGGTTGGATCCGATAGTG 1654

Qy  1702 caggcctctggacgacccgctcaagatcacgagttcacagacagggag 1761
Db  1655 AGAGGCTCTGCTGGGAGAGCCCACGGTGCAAGGTCCAAGTGATCATGAATGAAGAG 1714

Qy  1762 tcagcgaaaaggccttcttgtctccaacctcctgatctggctcaacaac 1821
Db  1715 CTCACGCTGGAGACACCACACTTGGACATGGAGCAGTGACATGAATGCCGATTTGC 1774

Qy  1822 ctgaaaagcggggagcacggtgcctgccaggctaccagtacagggagcttcagc 1881
Db  1775 TTGCAGAGGGCGCCGGGATACGATCTGGCAGCTGACGAGACAGCTTTGCTGGC 1834

Qy  1882 cagatgaaggctccggtgagctggaagaagacaccatggttttggtgaagaacatgtc 1941
Db  1835 TTGTCACGCTGGAGACACCACACTTGGACATGGAGCAGTGACATGAATGCCGATTTGC 1894

Qy  1942 gacaagatctgggactctgtacaaaatcctgacaacacatcgatgctgctgaggctta 2001
Db  1895 AACAAGATAATTGAGATTTCAGTATAAATTGAAGTTTCTTTGGCTGGTTCGG 1954

Qy  2002 gctggaaacttccccccggccgaaggaccccgctccggaaaggccggggtct 2061
Db  2002 gctggaaacttccccccggccgaaggaccccgctccggaaaggccggggtct 2061

Qy  2002 CTGGAAAGTTCTTCCTGCAGGTGGCACGCCGGACGCCGTGTCTATCGTTGGAAG 2014

Qy  2062 cagatggagcctccggagctctgcagtggagaacgacatacagatcaagcgg 2121
Db  2015 CAGATGAAGGCTCTCGTGAGGGATGGGGACAGGTTTTTCGGTCCCGGGACGACCTTCACA 2074

Qy  2122 gatgcagagaggcttgagctggagaacgacctcctctcggtcatcgtgaccttga 2181
Db  2075 GACGCTCAGAGGCTTGAGCTGGAGAACGAGTGCGTCACTTTGGCTCATCTCATCGTGACAGACTTGA 2134

Qy  2182 ctcctcacgactctgttctccatgcctcagaagacttgagctca 2241
Db  2135 GGACTCAGCACAGATTCTGTCGGATCCCTGTGCTCAACCTCTCCACTAGATAGT 2194

Qy  2242 tctgtacagaacatcagtgacaaaggctgacaggacctggacctcctccaggaac 2301
Db  2242 tctgtacagaacatcagtgacaaaggctgacaggacctggacctcctccaggaac 2301
```

```
db   2195  TCCTGCGAGACATCCCAGCATGCAACTGACCTACTATGGAGGACTTCCCACAGAT 2254
                                                                        RGSHAYVHPLVRRLMTDFQVMTELRPLQLQVDVFFRPWRELQEGSLQDFTVRGLLARPAK
db   2302  gacaagcgtggttccatcagcagagtggggagctctgtccaggtggagcctgtct 2361     LQVQLDWELELIRLPLYSNVTTLLALASLQAPQARGHGRIHLNEWREYCGSLRAELETP
db   2255  GACAAGTGTGGTTCCATCAGCAGAGTGGGGAGCTCTGTCCAGGTGGAGCCTGTCT 2314     AELKALIAMESMVNKIMELYKADNIDVMLGGLLAEPFLPRGARTPFLPRCIIGKGMAL
Qy   2362  gggagggcggcatctgtatcctgccgcgggatagagctccagagcacgcgacga 2421     RGGDRFWEMRGYNVFTDAQGQELEKRGLPLRVYCNMTGLTRVPYVPARFIGKFQDFESCE
db   2495  AGTCTTCAAGCTGCGCCCAGAAGCATCTGTTGTCTCTAGAGTCCTCGCGCACTCAA 2534     EIDSSMGKALRTPFDCHSNPYEPCNEVVRENDYFEPEEEPVATLGTVITGKFFSGQTGL
Qy   2422  ctcactgcacccagagagcaggatttcaccggcccctacaacagggctcgctat 2481     TCTGGNRNEQGVRMLETDDWRPELAPGCFPVLVTRTHADKKSTLLITERVPMESSPRK
db   2315  GCGAGACTGGTGTGTTGTGTTCATGTATGACATGGTCCAGACCGAGAGTCTC 2374     SQSVISKQKAKQYVAQGEAYGSRVLCE*
db   2495  AGTCTTCAAGCTGCGCCCAGAAGCATCTGTTGTCTCTAGAGTCCTCGCGCACTCAA 2554
db   2422  ctcactgcacccagagagcaggatttcaccggcccctacaacagggctcgctat 2481
```

BASE COUNT   690 a   772 c   728 g   513 t

ORIGIN

```
Query Match     46.9%;    Score 1441.8;  DB 10;    Length 2777;
Best Local Similarity  77.6%;
Matches 1791;  Conservative  0;  Mismatches 492;  Indels 24;  Gaps   3;

Qy    535  tacaggccccatacacggacagccaccccacccccaggatgaccgccgcccccaac  594
                 ||||| ||| ||||||||| ||| ||||||| ||||||||| ||| |||||||||||
db      1  TACCGGCCCATCACAGGGGTGTGCAACAACAGACATCACCCCAGATGGGAGCCTCCAAC   60

Qy    595  acgggcctggacagaatgtgctgctcgtatggcacgaagatggtttcagtcagccagaggc  654
                 ||||| |||| ||| |||||| ||| ||| ||||| || ||||||||| ||||||||||||
db     61  ACAGCCCTTACACAAGATGTGCTGCTCGTATGGCACGAAGATGGTTTCAGCCAGAGGC  120

Qy    655  tggaacccggctcttgtacaccggtcccacgccccggtccccgggagggtgacaaga  714
                 |||||||||||| ||||||||| ||||||||||||||| ||| ||||||||||||||
db    121  TGGAACCCGCTCTTGTACACCGGTCCCACGCCCCGGTCCCGGGAGGGTGACAGAG  180

Qy    715  catgcattcaagttccaacagaataatttgcaggatcacatctggctattcgacctccg  774
                 ||||| |||| ||||||||||| ||||||||||||| |||||||||||| ||||| |||
db    181  CACTCAATTCAAGTTCCAACAGAATAATTTGCAGGATCACATCTGGCTATTCTTCTG  240

Qy    775  atgcatgcggaccatacgggaccccacggcccagcagcccacgcagcagcagcacaa  834
                 ||||||||| ||||||| ||| ||||||| ||||||||| ||| || ||||| ||||
db    241  CCGGTGTGGCACATACTCGGGACCCACGGCCCAGCAGCCCACGCAGCAGCACA  300

Qy    835  gctgcttcgggaggggaagatgctgctgacagactgtccgtcacccctcccagcatgtttt  894
                 ||||| ||| ||||||||| ||| |||||||| ||| |||| || |||||||||||||
db    301  GCCAGACTTCTCGGGAGGGGTGACTGCCTTCCCGTCACCTCTCCCAGCATGCATG  360

Qy    895  cccataacaccctccgggagagcgagaggtgcgccgccgccgcgaccgcctcgccctcctac  954
                 ||||||||||||| ||||||||| ||| ||||| ||||||||||| ||||||||| |||
db    361  CCCATACAGTCCTCCAAACTCCTCACGACCGT----CACTGCATTCCCGTCACTTCTAC  414

Qy    955  cgtcttcggacgcggctggcgcggacgagggtcggcactgcttgggaaccctgtcga  1014
                 ||||||||| ||||||||||||||||||| ||||||||||| ||||||||||||||
db    415  CGTCTCTCAGCCGCCTGTGGCACTGGGAACCAGGACACTTCTTTGGCAACTGTCTGCA  474

Qy   1015  gccaacccggagcccagacagatgagggtttgacctcgttcctccacccccacctgtat  1074
                 ||||||||||||||||| ||||| || ||||||| |||| || ||||||| ||| |||
db    475  GCCAACCCGGAGCACAGTACAATGAGTGCTGTTCGAGAGAATGACTACTTGA  534

Qy   1075  gggagttcccacgacccgacggcaggagccgagacggccgaaagggaggcagcgctc  1134
                 ||||||||| || || ||||| ||| || || ||||||||| || || ||||||||
db    535  GGCCAAGTTCCCAAGTGCTGGAGAAGCAGTCCCGCACTGAAGTTCGCCAAGGTCCTG  594

Qy   1135  acccagcagcggcccccaaggcccactcgcaggagccccaggacggccgctcgatg  1194
                 ||| || || || ||||||| ||||| ||| || || ||||||||| || ||| ||
db    595  CGTGTCAAGCACCCTGCACCATCGTCCTGCCGCCCTGCCTGACTTGCCTCCTTT----C  642

Qy   1195  gcgcctcgagcctgtcgcgccgcaggggacctgcggcgccccctgctc  1254
                 ||| ||||| || ||||| |||| || ||| || || || |||| |||
db    643  GCATCAGCCCTGTGTCCTCCACAAGCCTGCTGTCCCTGGCCTCCTCGTAC  702

Qy   1255  ctgccggagacggccgcccgaggtccccctgaccggcactgaccagctgcagtatg  1314
                 ||||| ||||||||| |||| |||||||||| ||||||||| || ||| |||||||
db    703  CTGGCTGGTCGATGGGAGTGCAGAGGCATGGATGGCCTGTCTCGACAGTTGAGG  762

Qy   1315  ctcgggagctgaagccacctacttggtccgtcccggccaatcagccccaactgatg  1374
                 || ||||| || || |||||||| || || ||| || || |||| || || || ||
db    763  CTCGGGAGTTGAAGCCATCAACTGGTCCGTCCCGCCCAATCAGCCCCAACTGATG  822

Qy   1375  gacgcctgaccccgagagccgagcccgcaaggtcgtgggggtcctcgacatcatcaccgg  1434
                 || |||| || || ||||| || ||||||||| ||||||||| |||| ||| ||||||
db    823  GACGCCTGACCCGAGAGCGAGCCCGCAAGGTCGTGGGGGTCCTGATCATCACCG  822
```

RESULT 11

LOCUS   RATTPOPR   2777 bp   mRNA   ROD   27-APR-1993
DEFINITION Rat thyroid peroxidase (TPO) mRNA, 3' end.
ACCESSION   M31155
VERSION   M31155.1 GI:207434
KEYWORDS   thyroid peroxidase.
SOURCE   Rat thyroid cell line FRTL-5, cDNA to mRNA.
  ORGANISM *Rattus norvegicus*
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 2777)
  AUTHORS   Isozaki,O., Kohn,L.D., Kozak,C.A. and Kimura,S.
  TITLE   Thyroid peroxidase: rat cDNA sequence, chromosomal localization in
            mouse, and regulation of gene expression by comparison to
            thyroglobulin in rat FRTL-5 cells
  JOURNAL   Mol. Endocrinol. 3, 1681-1692 (1989)
  MEDLINE   90114171
FEATURES   Location/Qualifiers
     source   1..2777
                     /organism="*Rattus norvegicus*"
                     /db_xref="taxon:10116"
     CDS   <1..2313
                     /note="thyroid peroxidase"
                     /codon_start=1
                     /protein_id="AAA42265.1"
                     /db_xref="GI:207434"
                     /translation="VRPITCVCNNRDNRPHRGASNTRLARWLPPYTEDGFSPRGWRNN
                     FLQVGLDQYSRDGFHSMSREAVREDQSPRFLQWVWCGIMEDIALYSPDFSERSAAPSTA
                     PRQQMGDISFLDASYTVGSSPFGVLSREMFTRSSAACLRLRVTLHIILDGGMRTLPASRA
                     CAPRGAPHARNTPFCLAGCRASPYALAAVEDLAEDHGLATPFAFAITNRKSANF
                     ATGVRAWRVQALPGITWRDIYKILZPDFVCQVQEYPESTVNRPVMPYASTAAF"
```

```
RESULT 12
HPMTPOD
LOCUS       HPMTPOD      1045 bp    mRNA            PRI       14-JAN-1995
DEFINITION  Human thyroperoxidase mRNA, 3' end, and the Graves' disease
            mutation.
ACCESSION   M57702
VERSION     M57702.1  GI:339874
KEYWORDS    thyroperoxidase.
SOURCE      Human Graves' thyroid, cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1046)
  AUTHORS   Zanelli,E., Henry,M., Charvet,B. and Malthiery,Y.
  TITLE     Evidence for an alternate splicing in the thyroperoxidase messenger
            from patients with graves' disease
  JOURNAL   Biochem. Biophys. Res. Commun. 170 (2), 735-741 (1990)
  MEDLINE   90343792
FEATURES             Location/Qualifiers
     source          1..1046
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /dev_stage="Graves' disease patient"
                     /tissue_type="thyroid"
                     /map="2pter-p24"
                     /db_xref="GI:339874"
     gene            1..795
                     /gene="TPO"
     CDS             1..795
                     /partial
                     /gene="TPO"
                     /codon_start=1
                     /db_xref="GDB:G00-120-446"
     CDS             join(1..611,742..913)
                     /partial
                     /gene="TPO"
                     /note="Graves' disease mutated protein"
                     /codon_start=1
                     /db_xref="GDB:G00-120-446"
```

```
                    /product="thyroid peroxidase"
                    /protein_id="AAA61219.1"
                    /db_xref="GI:339875"
                    /translation="MWAWNSNVPTDAQRRRELEKSISLSRYICDRFGLTRVPRWDARQVGK
                    FFPDEEHLSSLLTNQFQIDSLSFLNRQDHRGDTLRQVKALFGLLAHVSRVLYPRLCSHRG
                    YRLGGNSLDGWRGLPRPVYFMSWLLLIGDPGFSTWVCNRRVLGNRKGLIQCR
                    ELGDDGRTVNCGPKHCDAHPCHAHGRRNGPKHKRFKGGFGCLQCAADFVP
                    ...
variation        611..742
                    /gene="TPO"
                    /note="TPO"
polyA_site       1046
                    /note="G00-120-446"
BASE COUNT   226 a   281 c   307 g   222 t
ORIGIN
```

Query Match 31.6%; Score 970; DB 9; Length 1046;
Best Local Similarity 100.0%; Pred No. 2.5e-167;
Matches 970; Mismatches 0; Indels 0; Gaps 0;

```
Qy  2092  tttggtgggaaaaacadccacacgtcttccacggatcgcaacagaggcgaggcagaactg  2151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  TTTTGGTGGGAAAAACAACCACACGTCTTCCACGGATCGCAACAGAGGCGAGGCAGAACTG  60

Qy  2152  tccctgtctcggtcatctgtacacacctggcctcaacaggtaccgatggacgcctc      2211
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  TCCCTGTCTCGGTCATCTGTACACACCTGGCCTCAACAGGTACCGATGGACGCCTC      120

Qy  2212  caagtcaggaagaactttggagctctgacggatcaccagggacaacctg          2271
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  CAAGTCAGGAAGAACTTTGGAGCTCTGACGGATCACCAGGGACAACCTG          180

Qy  2272  gaggccaggggaaacctttcctgcaagcaggaactgtggacgcacacaacgtgtggcttc  2331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  GAGGCCAGGGGAAACCTTTCCTGCAAGCAGGAACTGTGGACGCACACAACGTGTGGCTTC  240

Qy  2332  aatgggagctttgctgcacgtcggaggcgcgtgcgtgtgtctccgcgg          2391
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  AATGGGAGCTTTGCTGCACGTCGGAGGCGCGTGCGTGTGTCTCCGCGG          300

Qy  2392  cacgggtatcggatccaagccggaagcagtcggaacaggacattgggggg          2451
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  CACGGGTATCGGATCCAAGCCGGAAGCAGTCGGAACAGGACATTGGGGGG          360

Qy  2452  cagcctccccctcgcaaagtggacgatggcaagatggatggatttc          2511
          |||||||||||||||||||||||||||||||||||||||||||||
Db   361  CAGCCTCCCCCTCGCAAAGTGGACGATGGCAAGATGGATGGATTTC          420

Qy  2512  gcctctgcggggtcggcaagaaacacaaggcggctccacggacccctac          2571
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  GCCTCTGCGGGGTCGGCAAGAAACACAAGGCGGCTCCACGGACCCCTAC          480

Qy  2572  gagtcaggacgctggaggagaaggcctcgcctcccctcggggtgaggcac          2631
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  GAGTCAGGACGCTGGAGGAGAAGGCCTCGCCTCCCCTCGGGGTGAGGCAC          540

Qy  2632  atctccatgcaggctggcctgctcagctctgaagccatcctgagacaggcga          2691
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  ATCTCCATGCAGGCTGGCCTGCTCAGCTCTGAAGCCATCCTGAGACAGGCGA          600

Qy  2692  atttgccagggtgacaccactgcaccatctcggagcaagggga          2751
          |||||||||||||||||||||||||||||||||||||||||||
Db   601  ATTTGCCAGGGTGACACCACTGCACCATCTCGGAGCAAGGGGA          660

Qy  2752  gaactcccagcctcagagatcggggaaaccgggacgycctggggacccyttcgccc      2811
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  GAACTCCCAGCCTCAGAGATCGGGGAAACCGGGACGCCCTGGGGACCCTTCGCCC      720

Qy  2812  gcagctcagatccgggaagghagtctgaagaggaagggccggggatactacagagtcg  2871
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  GCAGCTCAGATCCGGGAAGGCAGTCTGAAGAGGAAGGGCCGGGGATACTACAGAGTCG      780

Qy  2872  ccgagggacctcgagagcaaagtcgagagcaactgagaacactcatgttcccaaaa  2931
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  CCGAGAGACCTCGAGAGCAAAGTCGCAGGACACTGCAGAACAGCTTCATGTTCCCAAAA      840

Qy  2932  tcaccgtcagactccttccaaacacagcgaaactgagaaatccagggccgagcgttttc  2991
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  TCACCGTACACACCAGCAAACACAGCGAAACTGAGAAATCCGGGCCGAGCGTTTTC      900

Qy  2992  caacaggggaaaactacagctactccatctgaggaggtgttggataaaagtg  3051
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  CCAACAGCGGGTAAATCTAGTTACCCATCTGAGGAGGTGTTGGATAAAATGT      960

Qy  3052  tatagctgca  3061
          ||||||||||
Db   961  TATAGCTGCA  970
```

RESULT 13

```
LOCUS       HDMTP007                1996 bp    DNA             PRI       25-APR-1996
DEFINITION  human thyroid peroxidase (TPO) gene, exon 8.
ACCESSION   M27707 J03856
VERSION     M27707.1  GI:339855
KEYWORDS    thyroid peroxidase.
SEGMENT     15 of 15
SOURCE      human lymphocyte DNA.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1996)
AUTHORS     Kimura,T., Ohtaki,S. and Kikkawa,F.
TITLE       Structure of the human thyroid peroxidase gene: comparison and
            relationship to the human myeloperoxidase gene
JOURNAL     Biochemistry 28 (10), 4481-4489 (1989)
MEDLINE     89352509
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by S.Kimura 23-JUN-1989.
FEATURES             Location/Qualifiers
     source          1..1996
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2pter-p24"
     exon            <1..680
                     /gene="TPO"
                     /note="TPO"
     prim_transcript <1..>1996
                     /gene="TPO"
                     /note="thyroid peroxidase mRNA and intron G"
     exon            681..1199
                     /gene="TPO"
                     /note="TPO"
                     /number=8
     intron          1200..>1996
                     /gene="TPO"
                     /note="thyroid peroxidase intron H"
BASE COUNT      377 a    599 c    645 g    375 t
ORIGIN
```

Query Match 16.8%; Score 516; DB 9; Length 1996;
Best Local Similarity 98.1%; Pred No. 1.6e-84;
Matches 522; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy   893  ttccctacacctcccgagggagcgcgccgcccgcgggaccgcctgttcgccctcctt  952
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   670  TTTGCCTGCAGTCCCGGAGAAGGAGCCCGGCGGCGACCGCCTGTTCGCCCTCCTT  729

Qy   953  accgctcctcggaccgggcgcgggacggcccaggcgcgtcctcttgggaacctgtcca  1012
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   730  ACCGCTCCTCGGCCGGCGCGCGGCGGGCACCAAGGCGCGCTCCTTCTTGGGAACCTGTCCA  789
```

```
Qy    1013  cgccaaccgcggagagaagaacggttaaccttcgggagcgtccaccggttccacccgtgt  1072
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     790  cggccaaccccggcgacagagatgaaccggtgacccgtttcctgcctgggacccgtgt    849

Qy    1073  atgcagtccccgccggccctgtgtggccgcggagctcgggaagctggagcgtggccatgctggtcgtgc  1132
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     850  atggcagtgctccccgcgcccctgtgggccagctggccagctggccaggggctgc    909

Qy    1133  tccgcgtccacgcggcctcccgggactccgggcgcctacctcgcgtgccgtcaccac  1192
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     910  tccgcgtccacgcggcgcctcccgggactcacgggcgctacctgcctgcgtgcgcac    969

Qy    1193  gcgcctctctgctcctggccctggccgggcagccccggagacccgggaagccgccctct  1252
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     970  gcgcctctcttgctcctggcctgggccggcagtcccggggacatccccgggagcccctct  1029

Qy    1253  tcctgcgcgagagccgccagccccggcccgagccgaggagaagaccccgggaacacccgtgt  1312
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1030  tcctgcgcgagagccggccagccccggcccgagccgaggagaagaccccgggaacacccgtgt  1089

Qy    1313  ggctcgcggacaaccgtcggcgggctcaaggccctcaatgacgacctcacgacg  1372
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1090  gctcgcggacaaccgccgcggggctcaaggccctcaatgacctcacgacgacc  1149

Qy    1373  cggagccgtgtatcacaaggaaggccctcgcaaggtcgggctcccgaccgcccgat  1424
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1150  cggagccggtgtaccaagaagaccccaagctgtgtggctgccacacacacagtt  1201
```

```
RESULT 14
ACC0G811
LOCUS       ACC0G811      160674 bp     DNA              HTC            04-MAY-2001
DEFINITION  Homo sapiens chromosome 2 clone RP11-112Bl1 map 2, WORKING DRAFT
            SEQUENCE, 15 unordered pieces.
ACCESSION   ACC0G811
VERSION     ACC0G811.3  GI:8783741
KEYWORDS    HTC; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 160674)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 160674)
AUTHORS     Anderson,S., Baldwin,J., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Bloom,T., Borowsky,M., Bastien,V., Beds,F.,
            Boguslavsky,L., Bouthgaiter,R., Brown,A., Burkett,G.,
            Champagne,M., Chapman,A., Chopart,T., Colmange,S., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Soyle,W., Ferreira,M., Filchak,M., Gage,D.,
            Gilbert,J., Gnerre,S., Gnirke,A., Goyette,M., Graham,J.,
            Grand-Pierre,N., Grant,D., Haas,B., Hesford,A., Horton,L.,
            Howland,J., Illiev,J., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,C., Lehoczky,J., Lewis,L., Liu,X., Locke,K., MacDonald,P.,
            Levine,R., Lieu,C., Liu,C., Locke,K., MacDonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McKernan,K., McPeeters,A., McPheeters,D.,
            Meldrim,J., Mesirov,J., Miranda,C., Morris,W., Mihalev,A., Morrow,J.,
            Murphy,T., Naylor,J., Morman,C.H., O'Connor,J., O'Donnell,P.,
            O'Neill,G., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pikhel,G., Rachupka,A., Raymond,C., Rogov,P., Rothberg,J., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,T., Vo,A., Wilson,R., Wu,X., Wyman,D., Ye,W.J.,
            Vassiliev,H., Viel,R., Viel,R., Volfovsky,N., Wyman,D.,
            Young,G., Zainoun,J., Zimmer,A., and Zody,M.
JOURNAL     Submitted (20-APR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 28, 2000 this sequence version replaced gi:8516129.
            All repeats were identified using RepeatMasker
```

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

```
            ---------------------- Project Information ----------------------
            Center project name: 1986D3
            Center clone name: 112_B_11
            ---------------------- Summary Statistics ----------------------
            Sequencing vector: M13; M13; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 156805 bases at least Q40
            Consensus quality: 156805 bases at least Q30
            Consensus quality: 158132 bases at least Q30
            Consensus quality: 159000 bases at least Q20
            Insert size: 159374; sum-of-contigs
            Quality coverage: 4.5 in Q20 bases; agarose-fp
            Quality coverage: 4.7 in Q20 bases; sum-of-contigs

*  NOTE: This is a 'working draft' sequence. It currently
*  consists of 15 contigs. The true order of the pieces
*  is not known and their order in this sequence record is
*  arbitrary. Gaps between the contigs are represented as
*  runs of N, but the exact sizes of the gaps are unknown.
*  This record will be updated with the finished sequence
*  as soon as it is available and the accession number will
*  be preserved.
            1886: contig of 1886 bp in length
       1887 1986: gap of 100 bp
       1987 4580: contig of 2593 bp in length
       4580 4679: gap of 100 bp
       4680  8193: contig of 3514 bp in length
       8194 8293: gap of 100 bp
       8294 12006: contig of 3713 bp in length
      12007 12106: gap of 100 bp
      12107 16433: contig of 4326 bp in length
      16433 16532: gap of 100 bp
      16533 21956: contig of 5424 bp in length
      21957 22056: gap of 100 bp
      22057 28704: contig of 6648 bp in length
      28705 28804: gap of 100 bp
      28804 36083: contig of 6893 bp in length
      36084 36183: gap of 100 bp
      36184 47720: contig of 11537 bp in length
      47721 47820: gap of 100 bp
      47821 58804: contig of 10984 bp in length
      58805 58904: gap of 100 bp
      58905 73984: contig of 15005 bp in length
      73985 73984: gap of 100 bp
      73985 89099: contig of 15005 bp in length
      89100 89099: gap of 100 bp
      89099 104855: contig of 15766 bp in length
     104855 104955: gap of 100 bp
     104955 125607: contig of 100 bp
     125507 125606: gap of 100 bp
     125607 160674: contig of 35068 bp in length
```

```
FEATURES             Location/Qualifiers
     source          1..160674
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-112Bl1"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1886
                     /note="assembly_fragment"
     misc_feature    1987
                     /note="assembly_fragment"
     misc_feature    4680..8193
                     /note="assembly_fragment"
```

```
LOCUS       SSTPORI      1142 bp    mRNA              MAM        13-JUL-1995
DEFINITION  Porcine mRNA for thyroid peroxidase C-terminal region (TPO).
ACCESSION   X04645 J03463
VERSION     X04645.1  GI:12141
KEYWORDS    thyroid peroxidase.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 1142)
  AUTHORS   Magnusson,R.P., Gaudin,M., Taurog,A. and Rapoport,B.
  TITLE     Isolation and characterization of a cDNA clone for porcine thyroid
            peroxidase
  JOURNAL   Mol. Endocrinol. 208 (2), 391-396 (1986)
  MEDLINE   87054617
COMMENT     Data kindly reviewed (23-MAR-1987) by R.P. Magnusson.
FEATURES             Location/Qualifiers
     source          1..1142
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /cell_type="thyroid cells"
                     <1..999
                     /codon_start=1
                     /product="thyroid peroxidase (332 AA)"
                     /protein_id="CAA28306.1"
                     /db_xref="GI:12142"
                     /db_xref="SWISS-PROT:P09933"
                     /translation="PCGLSRLETWADLSAVTANGKVADRILGLYGHPONIDVWIGGLA
                     ESFIPGAHTPLPACIIGKGHRALRLEDWNPVFTAQGRELSRHSSNSYIYIN
                     SGALAAFKWGNNPSGDVNCSRVDKRLESQVRHQEHDQGDVLIC
                     EERGQMVLVFSCHRGFRLRGPAQITCTPRGHDSPFLCKDINECRDTDPCHASARC
                     KNTKGSYLCSCSDPVVGEDGRTCVAGRELPRASIVSIALALRVAUVCGALAGTAVVCR
                     WTRAHANLLPOVGHDKSLPLPDCSHNRDVGAARALEYPDGLGCGSCHGLCE"
     CDS             1123..1128
                     /note="polyA signal"
     polyA_signal    177 a
                     /note="pot. polyA signal"
     polyA_site      363 c    408 g     194 t
                     /note="polyA site"
BASE COUNT      177 a    363 c    408 g     194 t
ORIGIN
```

```
Query Match       16.6%;  Score 510.4;  DB 4;  Length 1142;
Best Local Similarity   73.8%;  Pred. No. 2e 83;
Matches  649;  Conservative  0;  Mismatches 231;  Indels  0;  Gaps    0;

Qy  1873  tcctcgcgctcgtctcgggagacccccgtgacctgaacagacagcccgacggcagg 1932
          |||||||||| | |||||||| ||||| ||| |||||||||||||||||||||||||
Db 50676  TTCGCGCTCTCCCAGACTCGAGARCCCTTCGGCCCACCTGGTGCACCTGGTCCAACGAG  60

Qy  1933  aacgtgaccaacggctcgctgaccgactgagcaggacctgtcgctcgtcgtctcggctg 1992
          |||||| ||||| ||||||| |||| || |||||||||| | || |||||||||||| |
Db    61  CGTGTGGCGACGAATCCGGCCGCTCCAACCAGCACCCGCGCTGCTCGCGCCTGCCTG 120

Qy  1993  gagcttagctgaaaggctttcccccaggcctggaagggaccgccttcgccctg 2052
          || |||| || ||||||||| |||||||||||||||||||||||||||||||||
Db   121  GCGCGCTGGCCGAGACGTTCTTCCCTGGGCGGCGGGACGACCGCCTGCCGCCGCATC 180

Qy  2053  atcggaagcagatcaaaggctcgggaagcgcgtgactcgttctgttgggaggacggccac 2112
          ||||||| ||||| || |||| |||||| ||||| ||||| |||| |||||||||||||||
Db   181  ATCGGAAAGCAGATGAGAGGCCGTGGAGGGGANGAACCCCGGGG 240

Qy  2113  gtcttcagcggcctccagcagtactcggagcctgagggaccactcctcgggctcatcgt 2172
          |||||| | || ||| ||| || ||| || |||||||||||||||||||||||||||||
Db   241  CTGTTCACAAAGCCCAAGAGGGCCAGTTATCTGTCCTCTGCTGATCCTGTC 300

Qy  2173  gacaaacctggcctcacaggtcgccatactcccggggcctcaagtcgcaaatccccga 2232
          || |||| ||||| |||| || || ||||| |||||||||||||||||||||||||||||
Db   301  GACAACAGCCGCGTCCGAGGTGCTTCAGGGCTACGGCTCGGGGGGACCCTGCTCGGGGGGAGGGAGGGCTAC 360

Qy  2233  gacttggctgttgacgatcctcaggcatggcatccagggtctccaccaagt 2292
          |||||||||||| |||| || |||||| ||||||||||||| |||||||||
Db   361  GAGTTCCAGACCGGTGTACCAGGAGCCGGCAGGCCCT 420
```

```
BASE COUNT   39338 a  39937 c  38154 g  42844 t   1401 others
ORIGIN

Query Match       16.8%;  Score 516;  DB 2;  Length 160674;
Best Local Similarity   16.9%;  Pred. No. 3.8e-85;
Matches   98;  Conservative  0;  Mismatches  10;  Indels  0;  Gaps    0;

Qy   893  ttccatacaatccccgggaggaagccgcccctgcgctcgcccttcctct 952
          ||||| | |||||||||||||||| ||||||||||||||||||||||||||
Db 50735  TTGCCCGACTCCCGGAGGAAGCCCGGCCGCGCGCGCCTCGTCTGGCGTCT 50735

Qy   953  accgtctcttgggcccggctgccagcgggcgctcttctggaaacctgcca 1012
          ||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db 50736  ACCGCCTTCGGACGCCGTCGGCCACCCGGGACCCCTCTTTGGCGAACCTGCA 50795

Qy  1013  cggccaacccgagcagcagatgacgggttggcctgcctccacctgt 1072
          ||||||||||||||||||| ||||| ||||| |||||||||||||||
Db 50796  CGGCCAACCCGAGCAGCAGATGAACTGTTCTCTGGAACCGTGACCGGT 50855

Qy  1073  atggggcctcccccgggcccgcccagggctccaggaacatccggtggccc 1132
          |||| ||||| |||||| ||||||| |||||||||||||||| |||||||||
Db 50856  ATGGCAGCCTGCTTCGGCCACCCAGGGCTGTCAGGTGCCCAAGGGCTGC 50915

Qy  1133  tccggtcaggcccctgacgcggcagcccagctggccggagaaccctggtt 1192
          |||||||||||||||| |||||| ||||||||||||||||||||||||||||
Db 50916  TCCGGTCATGGCCCCCTGACGCGCGGCCTGCTGGTGCCGCCAC 50975

Qy  1193  gcgcggccttggtcgtctgccggcgccgaagcccgggccctgacg 1252
          ||||||||||||||||||||||||||||||||||||||||||||||
Db 50976  GCGCTGGTTCGTCTCGCCGGTGCCGAAGCCCGGGCCCTGACG 51035

Qy  1253  tcctggccgaaccggcgccgacgaagccccccccgccgacgcgacgctg 1312
          |||||||||||||||||||||||||| ||||| |||||||||||||||||||
Db 51036  TCCTGGCCGGAACCGGCGCCGCCACCGAGCCTCCGGGCCAGGAACCGGGCT 51095

Qy  1313  gtctccggaaacccggcctgccgcccggcgtccaaggccctcaaggcctc 1372
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 51096  GGCTGGCGGAGCACACCCCGCTGCCCGCCGGCGTCCAAGGCCCTCAAGGTC 51155

Qy  1373  cggacgcgcctgtaccaggaggcgcgcagagctgggctctgggcgaccaga 1424
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 51156  CGGACGCGCGTGTACCAGGAGGGGCGCAGAGTCTCCGGGCTGCCTCACAGGT 51207
```

```
     misc_feature    8294..12006
                     /note="assembly_fragment"
                     clone_end:896
                     vector_side:left"
     misc_feature    12107..16432
                     /note="assembly_fragment"
     misc_feature    15533..21956
                     /note="assembly_fragment"
     misc_feature    22057..28704
                     /note="assembly_fragment"
     misc_feature    36184..47720
                     /note="assembly_fragment"
     misc_feature    47821..58804
                     /note="assembly_fragment"
     misc_feature    53905..78904
                     /note="assembly_fragment"
     misc_feature    73985..89989
                     /note="assembly_fragment"
     misc_feature    89090..104455
                     /note="assembly_fragment"
     misc_feature    125507..160674
                     /note="assembly_fragment"
```

RESULT 15
SSTPORI

```
Qy  2293 cctcagacgcacaagtgtggcttcccagagagcgtgagaatggggacttgatgcactgt 2352
          ||  ||  || |||||||||| ||  ||| |||  |||||| ||| |||||||||
Db   421 CCGTCAGGGGACGCGTGTGCTTCCCGGACCAGTGGAAGACGGGGGCTTCCTGCTCTGT 480

Qy  2353 gaggagtctcggagagcgctgctgatgtatcctcgccgcacggtcatggcctccaaggc 2412
          ||||||   ||||||||||||| |||||||||||||| |||||||||||||||||||
Db   481 GAGGAGCGTGGCCAGCGCGTGCTGGTGTTTTCCTCGTCACGGGTTCCGGCTCCCAGGA 540

Qy  2413 cgggagcagtcacttgcaccagagagatgggatttccagcctcccctctgcaaagat 2472
          |||| |||||||||| |||||| |||||| || | ||||||| |||||| ||||
Db   541 CCAGCGCAGATCACTGCACACCCCCAGAGGGTGGGACTCCCCGCCCCCCTCTGTAAAGAC 600

Qy  2473 gtgaacagatgtgcagagagtgcccaacccccgccacgcctctgcgaggtgcagaaac 2532
          ||| || ||| |||||| |||||| |||| ||||||| ||||||||||||  ||||
Db   601 ATCAACAGTGTGCAAGACGAGACAGACCCTCCTTGCCACGCGTCTGCCGGTGCAAGAAC 660

Qy  2533 accaaaggcggcttcccagtgtctctcgcggcgagaccctacgactagagacacatggagaa 2592
          ||||| || || ||  ||||||||||||||||||||| |||||| ||||||||||
Db   661 ACCAAGGGTGGCGTCCGTCCTGTGCGAGTGCTCGGACCTTCGTGCTCGGGGAGGAGGACGGCAGG 720

Qy  2593 acctgcgtagactccgggagcctccctcgggtgacttggatctccatgtcgctggctgct 2652
          ||||||| ||| ||| |||||| ||| ||| |||| |||||||| |||||
Db   721 ACCTGCGTGGATGCGGCGGGAGGCTCCCGGGCGTCTGTGTGGTCTCATCGCCGCTGGCCGCC 780

Qy  2653 ctgctgatcgaggctccgagtctctcacctgacgtgatcttgaggtggacagcgact 2712
          ||||  ||||| |||| || || |||||| |||| | |||||||||||||| ||||
Db   781 GTGCTCGTCGGCGTCGGCGTCGCAGGCCTCGCCTGAGACGGTGGTTGCAGGTGGACACACGCG 840

Qy  2713 gtcactaaatcacacgcccctctcagagacagcgcgag 2752
          ||  ||   | ||||| || ||||||||| ||
Db   841 GATGCCAGGCCCTTGCTGCCCGTCCCGGGGAGGGAGGAAGGAG 880


Search completed. November 26, 2001, 10:03:06
Job time: 2816 sec
```